

## Louisiana State University LSU Digital Commons

---

LSU Master's Theses

Graduate School

---

2006

# Language classification and manipulation in Romania and Moldova

Chase Fauchaux

*Louisiana State University and Agricultural and Mechanical College, [cfauch1@lsu.edu](mailto:cfauch1@lsu.edu)*

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool\\_theses](https://digitalcommons.lsu.edu/gradschool_theses)



Part of the [Linguistics Commons](#)

---

### Recommended Citation

Fauchaux, Chase, "Language classification and manipulation in Romania and Moldova" (2006). *LSU Master's Theses*. 1405.  
[https://digitalcommons.lsu.edu/gradschool\\_theses/1405](https://digitalcommons.lsu.edu/gradschool_theses/1405)

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).

**LANGUAGE CLASSIFICATION AND MANIPULATION  
IN ROMANIA AND MOLDOVA**

A Thesis

Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Master of Arts

in

The Interdepartmental Program in Linguistics

By  
Chase Fauchaux  
B.A., Tulane University, 2004  
August 2006

## TABLE OF CONTENTS

ABSTRACT .....	iii
1 INTRODUCTION: AIMS AND GOALS .....	1
2 LINGUISTIC CLASSIFICATION .....	3
2.1 Genealogical Classification .....	4
2.2 Language Convergence and Areal Classification .....	9
2.3 Linguistic Typology .....	16
3 THE BALKAN <i>SPRACHBUND</i> .....	26
3.1 The Languages .....	29
3.2 The Features .....	31
4 ROMANIAN .....	39
4.1 The Romance Language .....	40
4.2 The Differences .....	47
5 THE HISTORY OF ROMANIAN .....	58
5.1 The History of the Two Theories and the Political Stakes .....	62
5.2 The Theory of Continuity .....	64
5.3 The Theory of Migration from the South .....	69
5.4 The Final Verdict: Which Theory Is More Solid? .....	78
6 STANDARDIZATION (AND MANIPULATION) IN THE WAKE OF ROMANIAN NATIONALISM .....	81
6.1 Transylvania .....	83
6.2 Wallachia and Moldavia .....	89
7 THE SOVIET UNION AND THE CREATION OF THE MOLDOVAN LANGUAGE .....	96
7.1 Language Policy in the Soviet Union: From Roots Revival to Russification .....	97
7.2 The Construction of the Moldovan Language .....	108
8 CONCLUSION .....	121
REFERENCES .....	125
APPENDIX: ALPHABET AND TRANSLITERATION TABLES .....	131
VITA .....	133

## **ABSTRACT**

Linguistic classification is a scientific methodology for categorizing the languages of the world. However, the tools and methods of linguistic classification have been used to various degrees by political entities to further nationalistic agendas. This thesis assesses the role of linguistic classification in nationality and politics, and addresses the disjunct between true linguistic classification, based on genealogical, areal, and typological features, and language designation, based on political and nationalistic motives. This thesis uses the Romanian language as a vehicle for illustrating both linguistic classification methods as well as how these methods have been manipulated for non-linguistic reasons both internally by Romanian nationalists and externally by the Soviet Union in Moldova. The Romanian language is analyzed in terms of three major classification methods: genealogical, areal, and typological. The thesis supports the idea that linguistic classification, a scientific practice, cannot be fully reconciled with language designation, a practice which essentially serves non-scientific purposes, namely to confirm or deny political and cultural relationships between different demographic groups. While some of the methods of linguistic classification are used in language designation, it must remain independent of real-world affairs in order to maintain its own principles.

## 1 INTRODUCTION: AIMS AND GOALS

The goal of this thesis is to explore the phenomenon of language classification and its implications for the historic and modern concepts of nationality, ethnicity, and identity. More importantly, I will assess the role of linguistic classification in the realm of nationality and politics, and illustrate the disjunct between true linguistic classification and language designation based on political and nationalistic motives. I will demonstrate how language designation driven primarily by political passions and agendas fails to comport with scientific classification methods.

This thesis will focus on the historical and modern classification of the Romanian language and the Moldovan dialect, spoken in the northeast of Romania and in Moldova and the Ukraine. Romanian occupies a unique position with respect to areal features and its position as a Romance language, and it has a history of both internal and external manipulative policies in Romania and Moldova. I will trace the history of the language from various angles of classification: in terms of its genealogical links to the Romance languages, its status as a full member of the Balkan *Sprachbund*, its historic ties with the Slavic languages, and how modern theories of linguistic typology can shed light on the classification of Romanian and Moldovan as they are spoken in the world today.

The concept of linguistic allegiance will remain at the heart of this paper. While linguistic classification should be a neutral, scientific issue, language designation is usually not neutral, and I will demonstrate how various political entities have tried to use some of the concepts of linguistic classification as a means to a nationalistic end: the designation of the status of languages, typically in relation to others, which can have enormous political consequences.

Romanian is historically one of the rougher edges of the Italic family of languages, perhaps the clearest and most well-attested language family in history and in modern linguistic science. Its development is somewhat mysterious, and it never appeared in print until well into the 16<sup>th</sup> century. Romanian language and culture exist on the very fringe of the Romance world. Because of a lack of developmental evidence (compared to other Romance languages and their speakers) and the tumultuous history of the Balkans which by no means left the language or its speakers unaffected, various groups with political motives have tried to shape the allegiance of the language's speakers by placing the language into contending historical frameworks. This has been attempted and accomplished by means of linguistic engineering to different levels of success by different groups who sought to align the Romanian language and its people to one political entity or another.

One of the most prominent of these groups was the 19<sup>th</sup> century Transylvanian School, a group of literati who sought to bring Romanian "back" into the Romance family using literary and political clout; the other major group was the Soviet Union and its legion of linguists, historians, and scientists who attempted to win the national allegiance of the Moldovan people by placing "Moldavian" (as they called it) into the linguistic theater of Slavic languages, designating it as a full-fledged language, separate from Romanian. The debate over the designation continues today, hardly ever without its political undertones, as Romanian and Moldovan are spoken in an area where uniqueness of nationality and identity are of tremendous importance.

## 2 LINGUISTIC CLASSIFICATION

No human language is unique in all of its features. Every human language shares at least some of its features with others, based on various affinities that are the result of historical consequences or forces within the very nature of human language itself.

Language classification is a practice which recognizes these affinities and seeks to group together languages which share common elements. The history of language classification is almost as old as language itself. The very first time one human being encountered another who spoke a different language, the recognition of that difference amounted to at least some rudimentary level of language classification. Over time, more complicated methods and theories would form. Perhaps the first of these to emerge was “genealogical” classification. The Romans believed (and erroneously so) that Latin had descended from Greek. Interestingly, even at this time such classification had political motives, since Greek was the prestige language of culture and refinement, even as the Romans ruled the known world. But the true breakthrough in genealogical classification came in the eighteenth century, when scholars noticed affinities between the ancient languages Greek, Latin, and Sanskrit -- affinities which could not be ascribed to chance. Thus, the concept of the Indo-European language family was born, and with it what would become the modern science of historical linguistics.

Another method of linguistic classification is to look at languages in terms of “language areas,” certain places in the world where languages have converged and assimilated to one another, often accompanied by cultural fusion as well. The Balkan *Sprachbund* is perhaps the best attested of these areas, where a group of languages (Romanian, Greek, Albanian, Bulgarian, and Macedonian, and to a lesser extent, Serbo-

Croatian and Turkish) from four distinct branches of the Indo-European family (except for Turkish) share features in their structure, vocabulary, derivational morphology, and idiom. Such linguistic areas can sometimes obscure a language's genealogical classification to the casual observer, since languages in these areas of convergence often end up having closer ties with each other than to their sisters and cousins in their own language families.

The final type of classification is linguistic typology, which has strengthened the other two methods considerably, while constituting a different kind of classification in its own right. Typology is the classification and taxonomy of languages regardless of history or current status in the world. Typology groups languages by features alone, and is often the check on the reality of any sort of genealogical or areal classification; without any sort of typological affinity, such ties are merely superficial.

While these three methods of linguistic classification do not in any way compete with one another, they have different functions. The true classification of a language will always be its genetic classification, but areal classification methods are used to demonstrate synchronic relationships with other languages outside of its linguistic family. Typology can be seen as a tool which allows linguists to strengthen claims about genetic and areal affinities.

## **2.1 Genealogical Classification**

As I mentioned previously, perhaps the oldest systematic method of language classification is the system of 'genetic' classification, which in the West can be traced back at least to the Middle Ages, as scholars, believing in the literal truth of the Biblical Flood story, intuitively arrived at the conclusion that all the world's languages must be



derived from those spoken by Noah's three sons. Some views tried to trace languages back to the Tower of Babel, and others attempted to cite Biblical Hebrew as the source of all modern languages. These narrow views lasted well into the early Modern Period, when such views gave way to more scientific ones based on empirical evidence. Several intellectuals noted the "family resemblance" among European languages, not only in lexicon, but also in certain morphological structures. Attempts at proving relationships between various European languages resulted in Sir William Jones' 1786 comment in "On the Hindus," in which he posits a single ancestor language not only for Greek and Latin, but also for Germanic, Celtic, Slavic, and most importantly, Sanskrit and Persian<sup>1</sup>:

The Sanscrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some source, which, perhaps, no longer exists; there is a similar reason, though not quite so forcible, for supposing that both the Gothick and the Celtick, though blended with a very different idiom, had the same origin with the Sanscrit; and the old Persian might be added to the same family.

This was the first step in recognizing the full scope of what would come to be known as the Indo-European language family. As these similarities could not be attributed to chance, the idea that these languages derived from a common ancestor and diverged due to rules of regular sound change became accepted, and in time other groups of languages were added; Slavic, Baltic, Albanian, Armenian, and even ancient Hittite were discovered to all have evolved from this common ancestor, which would become known as Proto-Indo-European. In fact, the concepts of evolution and "mutability" of the species were accepted by linguists long before biologists would take this theory as a given. In a short

---

<sup>1</sup> Quoted in Lass (1997), p. 106

time, the methods and principles espoused by the Neogrammarians were used to classify other languages of the world, as they are used even today.

Genealogical classification is a diachronic method based on the assumption that languages change over time, so that if one dialect of a language is separated from another for a significant period of time, assuming normal transmission (the transmission of an entire set of interrelated lexical and structural features)<sup>2</sup> they will dissimilate from each other as well as from the original language, resulting in new languages. This process is known as divergence, though many earlier theoreticians viewed it more negatively as a type of entropy or replication of error. A diagram representing this development of the original “parent” language over time would strongly resemble a family tree, or at least a cladistic branching diagram, similar to those seen in biology, except that the head of each branching node represents at least theoretically an actual language spoken at some point in time.<sup>3</sup> The mother language provides the roots, and the daughter languages make up the various branches as they differentiate from one another and from their “mother.” The smaller branches, which represent more recent divisions, may in turn break apart into still smaller branches. This *Stammbaum* model has lasted until the present day, and is commonly seen in texts dealing with historical linguistics.

Language change is by no means random, however, and at the most visible levels, phonology and morphology, it happens according to rules of regular sound change. This does not mean that it is predictable, but that we know what *can* happen and how it happens. In this way it can be likened to the weather: while the weather cannot be predicted, we know what is possible, and we also know how certain phenomena affect

---

<sup>2</sup> Thomason and Kaufman, p. 200

<sup>3</sup> This model was actually developed for historical linguistics and later borrowed by scholars in the fields of biology and evolutionary science.

others. Nineteenth-century Neogrammarians, such as Jacob Grimm, were able to perceive that cognate words among related languages, while varying from one another, do so in a highly systematic way. As this was recognized by the early predecessors of today's historical linguists, they realized they could use these rules as watermarks to date other rules, as well as to determine the relative relationship of one modern descendant to others with respect to the mother language. Of course, what naturally proceeds from this is the practice of reconstruction, which is an attempt at "recreating" the now-defunct mother language by comparing regular correspondences among its daughters. This Comparative Method is one of the oldest and most widely-trusted tools available to the historical linguist. While modern linguists do not try to make definitive claims about the actual status of the unattested proto-language, the mere acknowledgement of a systematic, synchronic relationship among a group of distinct languages which appears to result from normal transmission and not contact or chance will itself imply a diachronic relationship to some previous language; this realization alone is itself a reconstruction of the past.<sup>4</sup> According to Comrie, "a comparative reconstruction actually says nothing directly about history, but is rather simply a statement of the correspondences among a set of synchronic data."<sup>5</sup> Reconstruction is necessary in order to systematize cognate languages as they relate to each other synchronically. Only a few cases of reconstruction involve an actually attested parent language – within Indo-European, Latin and Sanskrit are the most readily available examples, but even then the spoken languages from which the modern ones derive were often far different from the language as it was attested in literary works which have survived. These attested proto-languages also bring to light the fact that there

---

<sup>4</sup> Comrie (2003), p. 246

<sup>5</sup> Comrie (2003), p. 247

will simply be some features which are missed, or some which result in the wrong conclusions. Assuming we had no knowledge of Latin, a reconstruction of Proto-Romance using the Comparative Method alone would yield results which we know to be erroneous because of our knowledge of the development of this language. For instance, we would never be able to reconstruct the final *//m//* of the accusative case, the case form from which most Romance nominals derive. For that matter, we would have no idea how to reconstruct the case system, or even be aware of its existence and extent, based on modern Romance data. Another famous example is the Latin future tense (*amo* ‘I love’, but *amabo* ‘I will love’), which would be absent from any reconstruction, as this morpheme which indicated the future tense in Latin has ceased to exist in its modern descendants.

The important lesson here is that reconstructions, while useful tools, can by no means be trusted as actual manifestations of the original language. They always remain theories, never to be taken as fact until an actual attestation is found. Comrie is also quick to note that any reconstruction is always open to replacement by a better reconstruction.

While genealogical classification has a high degree of charm and elegance, to believe in a strictly genealogical model of language change would constitute nothing more than a fantasy. It is a phylogenic practice, observing diachronic relationships among languages. However, diachronic studies hold no weight without synchronic correspondences. Based on further intuition from what is known about the synchronic relationships of languages in the world today, as well as further probing into history and the relationships that existed between languages over time, it is entirely obvious that languages do not exist within the confines of territorial boundaries. Even the most

consciously isolated language will exhibit evidence of contact no matter how hard its speakers (or, as is often the case, its politicians and grammarians) try to keep out such “impurities.” Language contact and convergence cannot be ruled out, and must be exhaustively considered in any theory of a language’s history, at all levels of analysis, from phonology to discourse pragmatics.

## **2.2 Language Convergence and Areal Classification**

Language divergence comes about in more or less two different ways: spontaneous change is the natural change over time in the language of a speech community as it drifts apart from the community of the original language. This process occurs from within, but change will just as inevitably occur from without. The latter is considered contact-induced change. As the various descendants of a language will diverge from one another, so too will they converge with the languages they come into contact with throughout their “lifetimes.” Language contact is indeed responsible for much of the dissimilation of dialects within a language which may otherwise be called spontaneous change, especially those changes which for one reason or another do not seem natural. When divergence occurs, it is usually more gradual than sudden, with continued contact with other varieties as well as outside languages often underlying the divergent developments. Usually, varieties continue to influence each other even as they move apart. Old English and Old Norse were relatively recently diverged Germanic languages when Norse had an impact on English in the Danelaw (Norse-occupied England). As a result, it is hard to tell sometimes when a word is borrowed from Norse or native to English. Only in extreme cases can a language be totally isolated from others, especially those which belong to its “family.”

Because of convergence and contact, it is possible to classify languages according to these relationships. Areal linguistics and classification involve the diffusion of structural features across linguistic boundaries. Such classification typically involves geographically contiguous areas, or at least historical geographic contacts (some languages which previously existed in contact with others may no longer share such contacts, such as the two branches of Finno-Ugric which eventually yielded Finnish in northern Europe and Hungarian much further to the south). Languages classified into these areal groups will share structural isoglosses which have arisen due to common contact and not common inheritance.<sup>6</sup>

The most common evidence of convergence can be seen in the very words I am writing now. Many of these words come from another source, and are not inherited from the Proto-Germanic cradle in which English was born and raised. Lexical borrowing, which includes the wholesale adopting of individual words as well as the translation of their constituent parts or ideas (calque or loan-translation), can be found in any language, and is the most common effect of language contact. Researchers sometimes have difficulty distinguishing loans from inherited words, especially when the two languages in question come from a common source.

Borrowing can be assumed to be internally motivated, whereby speakers of a language either unconsciously or consciously choose to adopt features of another language, for reasons which include (but are not limited to) necessity, where the borrower has no word for a certain concept (ex. English 'kangaroo' borrowed from the northeastern Australian language Guugu Yimidhirr), or perceived prestige of the donor language (English adoptions of French animal names for use in cuisine, such as 'pork,' 'beef,' and

---

<sup>6</sup> Bynon, p. 246

‘poultry’ instead of native English terms ‘pig,’ ‘cow,’ and ‘fowl’). Imposition, on the other hand, occurs when a political entity uses its power to overtly effect change in a language.<sup>7</sup> The efforts of the Soviet Union in the 20<sup>th</sup> century stand as testament to this phenomenon, as words, idiom, and structure were deliberately imported from Russian into the almost 200 minority languages of the country. The modern statuses of these languages can attest that such efforts were somewhat successful in some languages, such as Kazakh, but less so in others, such as the Baltic languages Latvian and Lithuanian.

Borrowing due to contact and convergence can manifest itself not only in the lexicon, but also in such perceived *sancta sanctorum* as idiom, syntax, semantics, and derivational morphology, and less commonly but by no means unheard-of, inflectional morphology. For example, in Romanian, the (now-dying) vocative case form for feminine nouns, a final //o// (nominative in //a// or //ə//), is borrowed from neighboring Slavic languages (neither Classical nor Vulgar Latin had a distinct vocative for feminine nouns). Campbell (1993) claims “given enough time and intensive contact, virtually anything can be borrowed.”<sup>8</sup> Core vocabulary, such as words involving basic kinship, daily life, and mundane natural phenomena and items, is considerably resistant to change, and therefore it is often used as evidence for genetic relationships. Morris Swadesh, in 1959 stated that “basic [core] vocabulary taken in sum is the one part of the language which most truly reflects the passage of time, and is least affected by special factors in the history of each group except that the specific changes may be influenced by the linguistic surroundings.”<sup>9</sup> While certain core items can be replaced (English pronouns *they* and *she* are from Norse, Spanish preposition *hasta* ‘up to; until’ is from Arabic) a

---

<sup>7</sup> van der Wurff, p. 384

<sup>8</sup> Campbell (1993), cited in Lass (1997), p. 189

<sup>9</sup> Swadesh (1959)

language which is separated from other members of its family by even thousands of years of outside contact will typically retain a significant amount of core vocabulary that matches the core vocabulary of its distant relatives. Armenian, one of the more isolated Indo-European languages, still retains many basic words which bear resemblance to forms in its distant cousin languages. For example: ‘cow’ – *kov* (Proto-Indo-European *\*g<sup>w</sup>ou-*); ‘work’ – *gorc* (PIE *\*wor-*), ‘father,’ Latin *pater* – *hayr* (PIE *\*pater-*); ‘foot,’ Greek *pod-* – *otn* (PIE *\*pod-m*); ‘sister,’ Latin *soror* – *k<sup>h</sup>oyr* (PIE *\*swesor-*); ‘eye,’ Latin *oculus* – *akn* (PIE *h<sub>3</sub>ek<sup>w</sup>*); ‘wet’ – *get* ‘river’ (PIE *\*wed-*); ‘ten,’ Latin *decem* – *tasn* (PIE *\*dek<sup>w</sup>m*).<sup>10</sup> Borrowing, even on a massive scale, can certainly never lead to a change in the genetic affiliation of a language, or even totally obscure it.

There are an infinite number of possibilities which may make up a contact situation. Oftentimes, when a new language moves into a region, it suppletes another, resulting in nothing more than trace features from the original language in the new adopted speech of the community, most often in phonology, but also in other areas, including core vocabulary. The Romanian word for ‘child,’ *copil*, is often attributed to this phenomenon, as it has no Romance or even other Indo-European counterparts. The converse situation involves an original language which is not supplanted, but is heavily influenced by the new language “from above.” Effects of this kind of contact often involve new lexicon, but do not typically change core vocabulary or inflectional morphology. The Old French influence on English is of this nature, whereby we get words like ‘person’ and ‘chair,’ as well as hundreds of words in the field of government, military, law, and urban life. The statuses of the languages involved in such relationships with one another are always uneven, which is why one language becomes dominant over

---

<sup>10</sup> Examples from Fortson, pp 340-349.



another. This imbalance is created by superior prestige, superior power, or superior technological advancement. The Spanish language, a language of colonization which is the offspring of a former language of colonization, Latin, demonstrates various examples of the aforementioned effects. For instance, the indigenous Iberian languages which would have been spoken throughout the Iberian Peninsula were supplanted by Latin, leaving a few trace elements which are debatably ascribed to the previous languages of the region, such as lexical items like *cama* ‘bed’ and *perro* ‘dog,’ which have no correlates in other Romance languages or in Celtic or Germanic, both of which were the source of many words in the ancient Romance world. On the other hand, as Spain remained under Umayyad dominion for over 700 years, the Arabic language has exhibited an enormous ‘top-down’ effect on Spanish, not only in lexical items but in phonology (some scholars point to the ‘j’ gaining its velar fricative quality) and in seemingly-closed word classes like prepositions (*hasta*, ‘until, up to’ from Arabic *ḥattā* of the same meaning) and perhaps pronouns (*usted*, 3ppl.-formal, possibly derived from Arabic *ustādh*, a term of respect which in the modern language means ‘teacher’ or ‘mister’ and not from the commonly attested *Vuestra Merced*, ‘Your Mercy’ -- this etymology is highly debateable and controversial).

In Latin America, the same effects can be seen, as Spanish acquired lexical items from Native American languages spoken throughout the Americas, especially in areas with a high Indian population (though the speakers themselves often have no proficiency in their ancestral language). Conversely, these languages bear a striking number of loanwords from Spanish, and not just in open word-categories.

The final contact situation is one where two or more languages exist in contact with one another but each has a more or less equal status relative to the others. While this is less common than the two previously mentioned relationship types, due to the overwhelming likelihood that languages and peoples in contact will exist in a dominant/subordinate relationship, it is a very dynamic situation in terms of linguistic contact, and is very important in the formation of the Romanian language. Such a contact situation invariably entails long-standing and widespread bi- and multi-lingualism, at least historically, and typically requires that all of the languages involved share similar levels of political import and prestige, lest one would oust the others. This type of relationship usually lasts longer than the other two, resulting in profound effects that typically extend much deeper than simple lexical borrowing. Also, when such a mutual exchange exists historically, it may be difficult to determine which feature originated where, especially when the languages in question also share genetic affiliation.

The result of this variety of prolonged contact can result in a *Sprachbund*, a term coined by Trubetzkoy and elaborated upon by Jakobson, who spoke of “linguistic alliances.”<sup>11</sup> A *Sprachbund* is a concentrated convergence area where otherwise unrelated languages bear salient similarities which could not have come about due to chance or by inheritance. Or, as Hock defines it:

Languages which may be quite distantly related or which exhibit no discernable genetic relationship may come to converge to the extent that they form a group that is structurally quite distinct from the surrounding and/or genetically related languages.<sup>12</sup>

The languages of the Balkans constitute the best-known *Sprachbund*, but others do exist, such as those of the Indian subcontinent, the Baltic region, Ethiopia, and

---

<sup>11</sup> Lehist, p. 59

<sup>12</sup> Hock, p. 494

Mesoamerica.<sup>13</sup> One of the interesting features of this effect is that it is often difficult to tell where a feature originated, or whether it originated from the languages in question at all. The postposed definite article, a tell-tale Balkan feature, is shared by all of the most central Balkan languages (except Greek, which has continued its historical preposing), yet its origin is unclear. Or, certain colloquialisms and expressions are common to all of the languages, yet where they came from, no one can say for sure. The Balkan *Sprachbund* will be further illustrated in subsequent sections.

Language contact often yields dramatic results over time. For instance, due to prolonged contact with France and the rest of Western Europe, English now has a vocabulary that is overwhelmingly French and Latin. In the 10,000 most frequently-used vocabulary items in English, 45% have French origins and 16.7% have Latin origins – only 31.8% of these words come directly from Old English.<sup>14</sup> English speakers may even have an easier time reading a passage in French than in Swedish for this very reason, even though Swedish is much closer to English in the genealogical sense. In fact, an English speaker may have an easier time reading French than he or she would have reading Old English! Facts such as these seriously confound the simple divergence model for language change as well as simple genetic classification. For example, unknown areal borrowings in the remote or not-so remote past can cause two genetically related languages to exhibit seemingly corresponding forms which may be ascribed to common inheritance and erroneously reconstructed as a feature of the parent language.<sup>15</sup>

So are languages really like cells which flagellate across time, occasionally dividing to form new cells? Or are they more like clouds, drifting apart and back into

---

<sup>13</sup> Campbell (in press), pp. 1-5

<sup>14</sup> Stockwell and Minkova, p. 50

<sup>15</sup> Campbell (1997), p. 54

each other, sometimes swallowed up, sometimes completely dissipating? Evidence of today's world favors the latter. And if what we observe is such a departure from the idea of clean breaks and divergence, then we can only assume that this has been the case as long as languages have existed. Contact and convergence have happened throughout history, at times when languages were not perceived as solitary entities as they are now. In reality, languages which derive from a common source probably remain in contact with one another, at least to some degree, continuing to share features and forms. Most of the modern Romance languages developed in such a way, where before standardization they existed in various shades of dialect which were by no means isolated from one another. Areal classification, while it by no means challenges genetic classification, shows that linguistic relationships are not as simple as a simple branching diagram. While the reality of a language's genetic classification can never be obscured, areal effects will force observers to recognize the genetic tree model as an idealization which does not necessarily indicate how "close" languages appear on the surface. Like a tree, while the leaves on one branch may encroach upon and grow into the leaves of another, they still belong to the original branch, no matter how well-mixed they appear to be with the other leaves.

### **2.3 Linguistic Typology**

Genetic and areal classifications are, in fact, taxonomic systems. They classify "living" languages in the world into various groups based in the former case on common inheritance and in the latter case on sharing of features within a geographical context. To take a zoological example, we know that wolves and lions are both mammals: they both have fur, they both have mammary glands, they both have upright legs, and they both

give live birth to their young, among other similar features. Of course, they have differences, but the fact remains that they share enough *core features* to be considered part of the same class: *mammalia*. Research into genetics and the fossil record can confirm that these animals (along with others) share a common ancestor millions of years in the past. We also know that not all mammals share the same features with the wolf and the lion: cows do not eat meat, apes walk on two legs, bats have wings, and dolphins must live in an environment which would kill a lion or wolf. On the other hand, there are features shared by lions and wolves which are shared by other animals that are not mammals: crocodiles are also carnivorous, birds are warm-blooded and care for their young, and hammerhead sharks give birth to live young. We can create classes such as “carnivores” which include any meat-eating animal, regardless of genetic affiliation, or “animals which give birth to live young” which would not include some mammals (monotremes, like the platypus), but would include non-mammals, such as the previously-mentioned sharks and some species of snakes. The point is, if we simply talk about features, then we can group species together and keep them apart regardless of whether or not they descend from a common ancestor. However, such feature grouping is also necessary if we were to try to prove any sort of genetic relationship, or if we were to make a statement about the features of all animals in a particular ecosystem (for instance, animals in Brazilian rain forests are often brightly colored, a feature not shared by animals living in the marshlands of Louisiana). What proceeds from this argument is that any sort of classification involves the identification of common features.

The purpose of this analogy is to illustrate the concept of linguistic typology, a more abstract form of linguistic classification than genetic or areal classification, but

necessary to validate either one of these classification methods. Lyle Campbell defines typology as “the classification of languages in terms of their structural characteristics.”<sup>16</sup> Typology is based on principles of language universals; the idea that there are certain features which must exist in language (very few, and extremely general, such as *all languages have consonants and vowels*), those which cannot exist (a voiced glottal stop consonant), those which can exist (any feature which can be found in a human language, such as grammatical gender or phonemic tone), and those which must exist if other features are present (*if a language has nasal vowels it must also have oral vowels*). The last type are called implicational universals, and are the ones which are most commonly dealt with in studies involving linguistic typology. While actual universals are rare on all counts (in all of the thousands of languages spoken in the present and in the past, one will almost always encounter exceptions), it is often possible to make generalizations so that we can say that one feature *usually* implies another, or that human language *usually* has or does not have a certain feature. In this way, it is possible to know which features or correlations of features are strongly preferred in human language, and which ones are marked and/or “deviant.” For example, consonants produced using a pulmonic airstream mechanism are found universally, and may therefore be called unmarked, whereas consonants which use a velar airstream mechanism, the “click sounds,” are only found in a limited group of languages, and are therefore marked features. Research in typology provides the facts whereby generalizations about the nature of language can be made, by defining the traits languages share and marking the limits of language variation.<sup>17</sup>

---

<sup>16</sup> Campbell (1997), p. 49

<sup>17</sup> Campbell (1997), p. 49

Armed with a typology of features, linguists can cut across all lines of history and geography to classify language types. For instance, we can speak of languages which are “pro-drop,” meaning they can or must drop the pronoun in the subject position of an inflected verb. Languages as unrelated as Arabic and Czech fit into this class. We can speak of nominative-accusative systems in which the object of a transitive verb is the marked form, such as in English and in Finnish, or of ergative-absolutive languages in which the subject of the transitive verb is the marked form, as in Basque, Berber, Tagalog, and the Maya languages.

Early typological classification was based on the erroneous assumption that some languages were inherently better than others, usually ethnocentric theories which placed European languages at the top. However, some of the concepts developed by these early “typologists,” such as the founder of modern economics Adam Smith (who distinguished between “original” and “compounded” languages), are still important today, if only for the idea that languages can be classified in meaningful ways across all other boundaries. In the early nineteenth century, Friedrich Schegel and August Wilhelm Schegel came up with the analytic/synthetic distinction based on the theories of some of their immediate predecessors, such as their uncle, Johann Heinrich Schegel (in 1764):

Alle Sprachen in der Welt haben nicht mehr als zwey Mittel, wodurch sie den bey Wörtern vorkommenden Nebenbegriffe andeuten können. Sie thun solches entweder durch Flexionen, das ist, durch gewisse Veränderungen an dem Hauptworte selbst, oder sie brauchen andere Worte dazu, welche dem Hauptworte beygefügt werden, und deswegen **Hülfsörter** (voces auxiliaries) heissen.

[All languages in the world do not possess more than two means to express secondary concepts that co-occur with words. They do so either through inflection, i.e., certain modifications within the main word itself, or they require other words which are added to the main words, and which are therefore called auxiliary words.]<sup>18</sup>

---

<sup>18</sup> citation and translation in Koerner, pg. 3

When the Indo-European language family was “discovered” in the late eighteenth century, intellectuals such as the Schegels quickly realized that typological classification had serious ramifications in the realm of genetic classification. The Schegels and others, such as Wilhelm von Humboldt, who is credited with the division of language types into isolating, agglutinating, and inflecting types, thought that these types represented various degrees of sophistication, and of course placed the Indo-European languages among the most sophisticated. This also played into the idea that older languages such as Greek, Latin, and Sanskrit were more “perfect” and “pure” than their “degenerate” daughters, a view widely held through the nineteenth century. However, it was the tradition of von Humboldt which eventually led to the modern science of linguistic typology, as it emphasized linguistic diversity and classified languages according to morphological-syntactic type.<sup>19</sup> This kind of classification, with modifications such as the addition of terms like *analytic*, *synthetic* and *polysynthetic* by Edward Sapir,<sup>20</sup> along with a lessening of the ethnocentric attitudes of many of the early typologists, continued well into the twentieth century and are still used to some degree today.

This sort of typology was often seen as mere taxonomy, and typologists of the twentieth century focused their energies on explaining linguistic phenomena and discovering general traits that may be found in many languages, particularly those with no common history. Roman Jakobson in 1958 claimed that typological

---

<sup>19</sup> Koerner, p. 4

<sup>20</sup> Koerner, p. 15



analysis could be used to make statements of nearly universal implication, which could be used to limit the scope of what is possible in language. This proved to be enormously important to the field of historical linguistics, since it provided a scientific basis to restrict the possibilities of linguistic reconstruction.<sup>21</sup> Joseph Greenberg greatly expanded these ideas with his system of “implicational universals,” usually in the form “If a language has X characteristic, it will (or will overwhelmingly likely) have Y characteristic.” Greenberg is also responsible for the classification of sentence types based on the relative positions of the subject, verb, and object – languages could be classified as SVO, SOV, VSO, VOS, etc., each type containing a whole host of likely implications about other grammatical matters (such as “languages with dominant VSO order are always prepositional” and “if a language has dominant SOV order and the genitive follows the governing noun, the the adjective likewise follows the noun”).<sup>22</sup>

Linguistic typology is important in the study of language universals because it uses diversity to posit uniformity in human language, giving us a clearer picture of what would or would not be a prototypical human language. In historical linguistics, it is not only important for the purpose of showing correspondences in structure, but it is for essential in reconstruction. Given a question of whether or not corresponding synchronic forms *d*, *d*, *t*, *t* in an intervocalic context should be reconstructed as *\*d* or *\*t*, there is no question that the voiceless variant should be chosen, since typological tendencies testify that acquisition of voicing in this context is much more likely than the loss of it.

Typology also keeps us from reconstructing forms that may at first seem to be apparent

---

<sup>21</sup> Koerner, p. 17

<sup>22</sup> Greenberg (1963) in Denning and Kemmer (1990), p. 45

but in fact go against the grain of human language, or from positing rules which make no sense ( $h > b / \_ \#$ ). It would be highly suspect to include a language in a family in which none of the members share any of the typological structural features of that language (i.e. adding a polysynthetic OSV language with glottal stops to Sino-Tibetan on the basis of seemingly similar vocabulary would be quite dubious). This is not to say that genetically related languages can not be different in typologies based on certain isomorphs. For instance, English is not a pro-drop language, but Spanish is, and most Indo-European languages are nominative-accusative languages, but some Indo-Iranian languages, such as Kurdish, have developed an ergative system due to contact with other languages within its geographical area, particularly ergative-absolutive Caucasian languages.

The function of the present discussion of linguistic typology is to assert its indispensability to genetic and areal classification systems. Synchronically, is a language family or a linguistic area really anything more than a collection of shared isofeatures? In a sense, such types of classification are typologies in context: the former context is one of historical development, the latter one of convergence and contact. By the same token, variation within these families can be based on a variation in type, so that one can claim that Spanish and Portuguese are separated from other Romance languages because they use a much more synthetic verbal inflection system than other languages within this family (French, Romanian, and Italian all typically use a periphrastic verb phrase for past tense whereas Spanish and Portuguese use an inflected preterite, for example).

The principles of linguistic typology are crucial to linguistic reconstruction. While Saussure claimed that genetic comparison without reconstruction is sterile by nature,<sup>23</sup> so too is reconstruction that does not consider language universals and general tendencies of

---

<sup>23</sup> Gamkrelidze, p. 25

human language and language change. For instance, the common reconstruction of the Proto-Indo-European oral stop system is voiced aspirate, voiced, and voiceless (example: \*/d<sup>h</sup>/, \*/d/, \*/t/). But this raises serious issues, since \*/b/ is in fact extremely rare in reconstructed forms. One must ask the question why it seems that the phoneme /b/, which is relatively unmarked and extremely common among languages of the world, is so uncommon while its counterparts of place (bilabials \*/b<sup>h</sup>/, /p/) and of manner (voiced non-aspirated stops \*/d/, \*/g/, \*/g<sup>w</sup>/, \*/g̃/) occur in more or less even distribution? Thomas Gamkrelidze claims that “Any linguistic reconstruction must be based on comparative evidence, and at the same time take into account the typological plausibility, both synchronic and diachronic, of a linguistic system arrived at by means of comparative and internal reconstruction.”<sup>24</sup> Or, as he states in the same article, “A linguistic reconstruction running counter to language universals cannot, naturally, claim to reflect a language that did historically exist.”<sup>25</sup> He then goes on to present the case for Glottalic Theory, a controversial theory which claims that the three-way Indo-European stop system in fact consisted of glottalized stops, aspirated/unaspirated voiced stops and aspirated / unaspirated voiceless stops. The claim is that in systems which do have a glottalized stop series, it is not uncommon for /p’/ to be rare or non-existent, whereas in a system which has a voiced stop series, /b/ would be quite common. This would also support the notion that the Indo-Europeans originally shared a homeland in contact with South Caucasian and Semitic languages – perhaps a glottalized stop series, itself a highly marked proposition, would be a result of prehistoric language contact. Such claims are hotly contested by other Indo-European scholars, but the important fact here is that we

---

<sup>24</sup> Gamkrelidze, p. 27

<sup>25</sup> Gamkrelidze, p. 27

cannot simply take the discrepancy in the consonantal system at face value. Because of what we know about language universals and what is natural in the typological sense, we know that such a system needs to be qualified and accounted for.

The modern discipline of linguistic typology, which has established universals as well as strong tendencies of language structure and development, provides linguists with the tools to test plausibility, as well as to be able to reject or suspect those reconstructions which go against them, and to establish credibility and even probability in those which adhere to them. One of the basic tenets of language typology which must be followed when venturing into any sort of diachronic study is the assumption that language has existed in its present state of complexity since long before the time period which is even theoretically possible to reconstruct, meaning that all typological universals and tendencies which are established by observing modern languages synchronically must also be true diachronically. An understanding of this fact is crucial to any sort of reconstruction, which cannot assume that the previous language existed in a more primitive or a more perfect state than it does now.

Typology is especially important when dealing with issues of language contact. Certain features can be shared by languages simply due to typological universals. For instance, both English and Mandarin Chinese are SVO languages. They also both share the features that adjectives precede nouns and that locative markers are prepositional and not postpositional. This should lead to no suspicion of contact, however, since both of these features naturally arise due to the sentence type. On the other hand, certain Bantu languages such as Zulu and Xhosa share highly marked phonological features with Koi-San languages in Southern Africa, most notably the famous clicks, or velaric airstream

ingressives. Since such features are in fact highly marked, one must assume that this feature arose due to contact and not to independent development. In a convergence area, a linguist must always ask the question as to whether a particular feature arising in two or more languages is necessarily due to contact or due to natural typological tendencies. For instance, there are certain features shared by Western Indo-European languages that are undoubtedly due to contact. There are others, such as the periphrastic past tense in both spoken French and German for instance (*j'ai mangé la pomme / ich habe den Apfel gegessen* where English could use the more synthetic simple preterite "I ate the apple," where "I have eaten the apple" has a specific meaning separate from the simple past tense). Could this have arisen because of universal tendencies of languages of this type, or was there some sort of areal contact phenomenon? The latter explanation is much more likely. The following sections will illustrate the role typology plays in establishing the Balkan *Sprachbund*, as well as in the placement of Romanian/Moldovan into both the *Sprachbund* and its own genetic family, the Romance languages.

### 3 THE BALKAN *SPRACHBUND*

The Balkan region of Europe is one of the most well-traveled crossroads in the world. Since ancient times, it has seen a host of invasions and migrations: Greeks, Romans, Slavs, Goths, Hungarians, Turks, and others have all made their marks there in one way or another. With such a high degree of cultural traffic, a linguistic and cultural melding can only be expected. The linguistic diversity in this area is a testament to the sheer number of different peoples who make or have once made this region their home. Speakers of Latin and its descendants Romanian (and various “dialects,” including Arumanian, Meglenitic, and Istro-Rumanian) and the now extinct Dalmatian, ancient and modern Greek, all of the South Slavic languages (Old Slavonic itself but also descendants Bulgarian, Macedonian, and Serbo-Croatian among others), Albanian, Turkish, Hungarian, the extinct Gothic language, and the ancient languages of Dacia, Thracia, and Illyria have all made their appearance in this region. Throughout this time, and especially in the 20<sup>th</sup> century, these groups have struggled to maintain their own separate identities. But amid the diversity, amid the struggle to stay unique, the Balkan cultures and the languages that serve them have achieved a remarkable degree of synthesis. The focus of this thesis, Romanian (and its dialect Moldovan) plays no small part in this linguistic melting pot.

Previously, I discussed the phenomenon of language contact and areal classification. Such a classification would involve the diffusion of structural features across linguistic boundaries. Any area in which the languages of the regions share these sorts of characteristics is called a *Sprachbund*, or a “linguistic union.” The Balkan region makes up what is perhaps the best known linguistic union, and is certainly the most

extensively studied. The Balkan languages share a number of features which can only be attributed to interlanguage contact, such as the famous postposition of the definite article, the loss of the infinitive verbal form, and the move towards more analytical verbal constructions (as opposed to the inflected, synthetic ones seen in their genetic relatives), to name but a few examples.

This Balkan phenomenon was first explicitly noticed in 1829 by the scholar Jernej Kopitar, who noticed that the languages spoken south of the Danube River all shared common formations, expressed by way of “different language material.”<sup>26</sup> In 1928, Nikolay Trubetzkoy first referred to this type of linguistic situation (specifically referring to the Balkan region) as “Sprachbund,” noting that they are remarkably similar in sentence structure and word formation, while showing no systemic sound correspondences in the way that genetically-related languages would.<sup>27</sup> Scholars have debated the origins of this “linguistic union” in the past. The earliest Balkan linguists, including Kopitar himself, asserted that the features making up the Sprachbund were somehow descended from the ancient Balkan substrate languages Dacian, Thracian, and Illyrian. However, this cannot be proven, since we know very little of these languages. Besides, the properties associated with the linguistic area developed during the Byzantine period, after the aforementioned substrata were already long dead. In the milestone *Linguistique Balkanique* (1930), Kristian Sanfeld proposes that these Balkanisms are due to the influence of ancient Greek, a language which, according to him, “malgré tous les decadences, il n’a pas cessé d’être porteur d’une civilisation supérieure à celle ses voisins” [despite all of its decadences, it never ceased to be a carrier of a civilization

---

<sup>26</sup> Tomić, p. 1

<sup>27</sup> Tomić, p. 1

superior to that of its neighbors].<sup>28</sup> This view was also shared by Alexandru Rosetti, a prominent historian of the Romanian language, in the 1965 *Istoria limbii române*. However, once again, the Balkanisms in the neighboring languages arose after the classical period, as did those Balkan elements in Greek itself, so influence from Ancient Greek is highly unlikely. Besides, Greek exhibits fewer Balkan features than do other languages within the linguistic union, such as Bulgarian and Romanian. However, Byzantine Greek could very well be the origin of many Balkanisms, due to the extensive influence of the Byzantine culture in the Balkans throughout the Middle Ages (Constantinople was the central city in the whole area at that time, and it exerted tremendous cultural and political influence).<sup>29</sup> As late as 1980, Georg Solta claimed that Latin was the source of them, though this is at best unlikely, since neither Latin nor any of the Romance languages that arose west of Romanian appear to show any of these features, or even any tendencies towards them.

The actual answer is that the properties of the Balkan Sprachbund come from the multilingual environment itself. That is, such a mixed situation may have arisen from the convergence of multiple languages within a limited area. In fact, some of the features can indeed be traced to a single source – certainly common words may be attributed to their source language (often Greek, Latin, or Turkish), and even some syntactic features, such as the analytic formation of the perfect tense, the form of which most likely comes from late Latin, and is evidenced in all of the Romance languages today (though to a lesser degree in the Iberian languages). This convergence model is accepted by most Balkan linguists as the most likely explanation of the shared features, and fits in with the

---

<sup>28</sup> Sanfeld, p. 17; Tomić, p.2

<sup>29</sup> Bynon, p. 246



previous discussion of linguistic areas and linguistic convergence. Tomić (2003) writes that “the convergence model is corroborated by the fact that the Balkan Sprachbund properties are most numerous in those parts of the Balkans where the greatest number of languages are co-territorial.”<sup>30</sup> This is true in that, historically, the languages which share the largest number of distinctly Balkan features (Romanian, Bulgarian, Macedonian, and Albanian) existed in very close proximity to one another (the issue of Romanian and its “migration” north will be discussed later), though they may not exist so close to one another in the modern era.

### **3.1 The Languages**

The Balkan Sprachbund involves at least seven modern standard languages: Albanian, Romanian, Bulgarian, and Macedonian are what may be called Balkan languages of the first grade.<sup>31</sup> Balkan languages of the second grade include Serbo-Croatian and Greek, while Turkish is more or less only on the periphery. In addition to these languages, one may also include several languages spoken by very small numbers of people, the most important one being Arumanian, a Romance language spoken in Greece, Albania, and Macedonia, which is closely related to modern Romanian. For the purposes of this paper, most of the examples given will come from Romanian, Bulgarian, Albanian, and Greek, as they are the best-documented in Balkan literature.

Auguste Schleicher characterized the Balkan language as “a group of languages, which have in common that they are the most corrupted members of their respective families,” reflecting the nineteenth-century bias which regarded Classical and Western

---

<sup>30</sup> Tomić, p. 2

<sup>31</sup> Du Nay, p. 86



**Figure 1.** Map of the Balkan area, with core Balkan Sprachbund areas in white. Peripheral areas are shown in light grey (map from the Regional Environmental Center with modifications).

European languages as superior.<sup>32</sup> However, the modern linguist is able to discern that membership in this Balkan “club” is based on shared features and shared typological properties. It is through objective linguistic typology that we are able to discuss characteristically Balkan features, even if we do not necessarily assert that existence of a Balkan-type classification. Tomić (2003) gives a set of criteria to follow when considering a feature as defining a language’s membership in a Sprachbund:<sup>33</sup>

- a.) the feature must be shared by at least three languages in the area, at least two of which must come from different genetic families
- b.) the features exhibited must not show up in other languages which are in the family but not considered as part of the Sprachbund.

She also goes on the claim that since the amount and extent of areal typological properties necessary for membership cannot be assessed independently, discussion on

<sup>32</sup> quoted in Bynon, p. 246

<sup>33</sup> Tomić, p. 4

Balkan membership has always focused on specific shared properties.<sup>34</sup> However, in reality, aren't all types of synchronic linguistic classification more or less based on the bunching of shared properties? Even diachronic classification, such as the genetic classification used in historical linguistics, can be corroborated or even proved or disproved by the existence or nonexistence of shared isolectal features.

### 3.2 The Features

While many features exist which can be called “Balkan features,” membership in the Sprachbund is characterized by at least one phonological feature – the existence of the mid-central lax “schwa” vowel ([ə], represented in Romanian and other Romance dialects as ‘ă’, in Albanian as ‘ë’, and in Bulgarian as ‘ъ’, which I have transcribed as ‘ă’ throughout), and the following morphosyntactic features (others do exist, but the following are the most prominent; there are also unconfirmed concordances which are still being investigated):<sup>35</sup>

1. a grammaticalized and postposed definite “article”
2. clitic doubling (use of both a clitic pronoun and its referent in the same sentence)
3. loss of the infinitive and substitution in most cases by the subjunctive
4. use of a bare subjunctive to express a mild command, a desire, or suggestion
5. merging of the dative and genitive cases
6. future tense expressed analytically (often using “want” as an auxiliary)
7. an analytic perfect marker using a “have” auxiliary
8. verbs that take two direct objects
9. loss of the distinction between question words and their non-interrogative counterparts
10. an analytic formation of the comparative for adjectives
11. a common pattern for constructing the numerals 11-19
12. an overall tendency to replace case endings with prepositions

While these morphosyntactic features are shared by all or most of these languages, it is always important to remember that the actual lexical and morphological material used

---

<sup>34</sup> Tomić, p. 4

<sup>35</sup> Tomić, p. 4

comes from native sources within each language (that is, the features are shared, not the morphemes themselves).

The postposed definite article is quite possibly the most well-known morphological feature of the Balkan Sprachbund. Note that Romanian is the only Romance language which has an enclitic definite article (though this is found in Germanic Scandinavian languages), and that articles are notably absent in Slavic languages besides Bulgarian, Macedonian, and the easternmost dialects of Serbian. Albanian has no living relatives in its sub-family, and Greek does not participate in this feature (it uses a free, preposed article: *o anthropos*, ‘the man’):

- a. Romanian (R): *lup*, *lupul* (‘wolf, the wolf’); *apă*, *apa* (‘water, the water’)
- b. Bulgarian (B): *măž*, *măžăt* (‘man, the man’); *žena*, *ženata* (‘woman, the woman’)
- c. Macedonian & Eastern Serbian (Torlak) (M & ST): *maž*, *mažot* (‘man, the man’); *žena*, *ženata* (‘woman, the woman’)
- d. Albanian (Al): *mik*, *miku* (‘friend, the friend’); *djal*, *djali* (‘boy, the boy’)

Clitic doubling is found throughout the Balkan Sprachbund, and is usually used in order to place, emphasis, focus, or topicality on a particular substantive. In Romanian, clitic doubling is obligatory when animate objects are involved (though all of the Balkan languages exhibit features such as this at least to some degree). Here is the sentence “I see George”:

- a. R: *Îl văd pe George*.  
‘Him see-I on[+acc](Accusative preposition) George.’
- b. B: *Viždam go Georgi*. (colloquial variant in SVO sentences, but obligatory in topicalized object sentences: *Georgi go viždam*.  
‘See-I him George.’ / (‘George him I-see.’)
- c. M: *Go gledam Gorgi*  
‘Him see-I George.’
- d. Greek (G): *Ton vlépō ton Giōrgo*.  
‘Him see-I the[acc.] George.’
- e. Al: *I shikoj Gjergjin*.  
‘Him see-I George-the[acc].’

The “loss of the infinitive”, though it sounds quite dramatic, has only come to full completion in some of the Balkan languages: Albanian, Macedonian, and Greek. In fact, this innovation is thought to have arisen from Greek, which was already replacing the infinitive with the subjunctive by the period in which the New Testament was written. In Bulgarian, the Romance dialect Arumanian, and many Serbian dialects, it is all but gone. In fact, this is one of the distinguishing features which separates the two “languages” Serbian and Croatian, which by most accounts are two highly mutually intelligible dialects separated for political reasons. And in modern Romania, the standard written language includes the infinitive, but the spoken language has ubiquitously replaced it with the subjunctive. This phenomenon is also observed in Turkish dialects spoken in Bulgaria, but not at all in standard Turkish. Here is the sentence “I want to write”, which is rendered more or less ‘I want that I write [+subjunctive]’:

- a. R: *Vreau să scriu*. (with the rarely-spoken infinitive, *Vreau a scrie*.)  
 ‘Want-I that[SC – subjunctive complementizer] write-I [subj.]’ /  
 (‘Want-I to write[inf.]’)
- b. B: *Iskam da piša*.  
 ‘Want-I that[SC] write-I.’
- c. M: *Sakam da pišuvam*.  
 ‘Want-I that[SC] write-I.’
- d. Serbian (S): *Želim da pišem*. (Though Croatian dialects use the infinitive, *Želim pisati*.)  
 ‘Want-I that[SC] write-I.’ / (‘I-want to-write.’)
- e. G: *Thélō na grápsō*.  
 ‘Want-I that[SC] write-I [subj.]’
- f. Al: *Dua të shkruaj*.  
 ‘Want-I that [SC] write-I [subj.]’
- g. Bulgarian Turkish (BT): *İsterim yazayım* (in standard Turkish: *Yazmak istiyorum*.)  
 ‘Want-I write-I[subj.]’ / (‘To-write I-want.’)

All of the Balkan languages have the ability and tendency to use a bare subjunctive verb to express a mild command, suggestion, desire, or intention. Here is “You should go!”:

- a. R: *Să te duci!*  
‘That[SC] you[reflexive] go-you[subj.]!’
- b. Arumanian (Ar): *S-ti duts!*  
‘That[SC] you[reflexive] go-you [subj.]!’
- c. B: *Da otideš!*  
‘That[SC] go-you!’
- d. M: *Da ideš!*  
‘That[SC] go-you!’
- e. ST: *Da ideš!*  
‘That[SC] go-you!’
- f. G: *Na pas!*  
‘That[SC] go-you [subj.]!’
- g. Al: *Të shkosh!*  
‘That[SC] go-you[subj.]!’

All of the Balkan languages at one time had distinct genitive and dative case nominal and pronominal forms. The modern languages, however, have merged both of these cases, using one morphological form to represent both. In Romanian, nominal case does exist, but is only marked on feminine singular nouns, so that *casă* ‘house’, yields *case* ‘of a house’ in the genitive / dative case. In other genders (and also in certain feminines), case is only marked on those forms which have a definite article attached (so the case-marking is more often than not technically on the definite article: *lupul* ‘the wolf,’ *lupului*, ‘of the wolf’). But Romanian has actually retained more of the Latin case system than any of the other Romance languages, which only account for case in the pronominal system. In Bulgarian and Macedonian, the dative and genitive forms are created by simply adding a preposition, which is the same for both genitive and dative constructs. The following are the sentences “I gave the book to Maria” and “It is Maria’s book” in various Balkan languages:

- a. R: *I-am dat careta Mariei.* / *Este cartea Mariei.*  
 ‘Her-have-I given book-the Maria[gen/dat].’ /  
 ‘Is book-the Maria[gen/dat].’
- b. B: *Dadoh knigata na Marija.* / *Knigata e na Marija.*  
 ‘Gave-I book-the on[gen/dat] Maria.’ /  
 ‘Book-the is on[gen/dat] Maria’
- c. M: *Ja dadov knigata na Marija.* / *Knigata e na Marija.*  
 ‘I gave-I book-the on[+gen/dat] Maria.’ /  
 ‘Book-the is on[+gen/dat] Maria.’
- d. Al: *Ia dhashë librin Marisë.* / *Është libri i Marisë.*  
 ‘Her gave-I book-the[acc] Maria-of[gen/dat].’ /  
 ‘Is book-the her[gen/dat] Maria-of[gen/dat].’
- e. G: *Edōsa to vivlío stē María.* (but the common colloquial variant is *Edōsa to vivlío tēs Marías*) / *Eínai to vivlío tēs Marías.*  
 ‘Gave-I the[acc] book[acc] on Maria[acc].’  
 (‘Gave-I the[acc] book[acc] the[gen/dat] Maria[gen/dat].’ /  
 ‘Is the book the[gen/dat] Maria[gen/dat].’)

The future tense in all of the Balkan languages is expressed analytically. They all follow the same pattern for doing so, probably once again modeled from Greek and its own development. This pattern uses the verb “to will, to want” followed by the verb, usually in the subjunctive. In most of the Balkan languages, the form originally meaning “to will” has evolved into an invariant form which marks the future tense (marked in the following examples as FC – future complementizer). Some, including Romanian, have kept the conjugated form in the standard literary language. Here is the sentence “I’ll see”:

- a. R: *O să văd.* (colloquial, with the literary variant *Voi vedea.*)  
 FC ‘that[SC] see-I[subj].’ (‘will-I see[inf].’)
- b. B: *Šte vidja.*  
 FC ‘see-I.’
- c. M: *Ke vidam.*  
 FC ‘see-I.’
- d. S: *Ja du da vidim.* (colloquial, with the literary variant *Ja du videti.*)  
 ‘I FC that[SC] see-I.’ (‘I FC see[inf].’)
- e. G: *Tha dō.*  
 ‘FC see-I[subj].’
- f. Al: *Do të shikoj.*  
 ‘FC that[SC] see-I[subj].’

All of the core Balkan languages form the numbers 11-19 using the paradigm ‘number on ten.’ This probably came about due to Slavic influence, as it is widespread in Eastern and Western Slavic languages as well as in the Southern languages spoken in the Balkans (Russian *odinnadcat’* (*odin-na-desāt’* ‘one-on-ten’), ‘eleven’; Czech *jedenáctka* (*jeden-na-deset*), ‘eleven’<sup>36</sup>). It is also absent in the Romance languages except for Romanian, as well as in Greek (Spanish *once*, French *onze* derived from Latin *undecim*, ‘one-ten’; Greek *éndeka* is formed the same way). Here is the word for ‘eleven’ in the Balkan languages:

- a. R: *unsprezece* (*unu-spre-zece*, where *spre* is a reflex of Latin *super*)
- b. B: *edinadeset* (*edin-na-deset*)
- c. M: *edinaeset* (*edin-na-deset*)
- d. S: *jednaest* (*jedan-na-deset*)
- e. Al: *njëmbëdhjetë* (*një-mbë-dhjetë*)

There are, of course, other shared features, as well as a vast amount of shared vocabulary and idiom. What follows are just a few of the idiomatic expressions shared throughout the Balkan Sprachbund: the word for ‘to ripen’ in Albanian, Greek, and Romanian is also the word for ‘to bake’. In many of these languages the expression ‘to kill oneself’ is formed by calquing the phrase ‘to remain without mouth,’ ‘to eat oneself with somebody’ means ‘to quarrel,’ and ‘to eat somebody’s ears’ means ‘to make a loud noise.’<sup>37</sup> ‘For many years’ is formed as ‘to many years,’ possibly based on the Latin *ad multos annos*:

- a. R: *la mulți ani*
- b. B: *za mnogo godini*
- c. M: *za mnogu godini*
- d. G: *eis pollá étē*
- e. Al: *për shumë mot*

<sup>36</sup> Note also that *c* typically represents the affricate [ts] in Slavic orthography.

<sup>37</sup> Bynon, p. 247



Finally, the expression ‘whether one (verb)s or not’ is rendered as ‘(verb)-not-(verb)’ (oddly enough, this is also how it is formed in Mandarin Chinese). Here is ‘whether one wants to or not’ (note that Turkish also participates in this syntactic feature):

- a. R: *vrea nu vrea*
- b. B: *shte – ne shte*
- c. S: *hteo – ne hteo*
- d. G: *thélei de thélei*
- e. T: *ister istemez*<sup>38</sup>

All of the languages also share morphological, syntactic, and idiomatic structures which can only be attributed to convergence. Sanfeld himself listed over one hundred concordances, though he was careful to note which ones were general concordances and which ones existed between one language and another.

The Balkan Sprachbund, as evidenced by the previous examples, serves as a testament to the power of linguistic convergence. Here we have a set of languages which can be classified into a definite group, yet we cannot construct a Balkan family, nor can we construct a Balkan type. As late as the 1930s, Alexandru Graur criticized the very idea of “Balkan linguistics,” saying that one can talk about “relationships of borrowings, of influences, but not about Balkan linguistics”.<sup>39</sup> However, it would be a grave error if the linguistic taxonomist were to overlook such glaring affinities. While few would argue with the statement “Romanian is related to Spanish,” would it follow that the statement “Romanian is related to Albanian” necessarily be wrong? Certainly, the shared history between Romanian and the other Balkan languages has been longer and closer than the shared history between the different dialects of Latin (after its diffusion across Europe),

---

<sup>38</sup> All linguistic examples in this section come from Tomić and the *Wikipedia* article “Balkan Linguistic Union”

<sup>39</sup> Du Nay, p. 87

which would eventually yield our modern Romance languages Romanian and Spanish. By the same token to say that since English is a Germanic language, it has more in common with Icelandic than French, would simply be wrong. In both of these cases, it seems that affiliation through contact has become stronger than affiliation through common inheritance, at least in the eyes of the speaker or casual observer. However, what is especially interesting is that if we look at these issues synchronically and diachronically together, the relationship of English to Icelandic and to French are not mutually exclusive, and English is “allowed” to have strong affinities with both. In actuality, the existence of “Balkan linguistics” or areal linguistics in general is no less valid than that of genetic or historical linguistics. No form of classification is clear-cut, and only by using synchronic typological methods to track areal features, as well as using diachronic methods of genetic classification (though convergence features themselves can also be looked at diachronically), can a linguist even attempt to make a scientifically sound categorization of a language. Even then, it is likely that the language in question belongs to several different groups and can be classified in several different ways, as Romanian is not only a Romance language, but also a Balkan language.

## 4 ROMANIAN

The Romanian language is genealogically a Romance language spoken by some 24-26 million people worldwide as the official standard in Romania and Moldova, where it is called “Moldovan” and considered a distinct language by some, in Hungary, the Ukraine, and in various parts of the Balkans, such as the Vojvodina autonomous region of Serbia and Montenegro. Due to relatively recent migrations, Romanian is also spoken by significant numbers of people in Russia, the United States, Canada, and Israel. The term “Romanian” may also be used to indicate a particular sub-family of the Eastern Romance languages, which includes Aroumanian (spoken in isolated pockets mainly in Greece, Albania, and Macedonia), Istro-Romanian (spoken by a small number of people on the Adriatic coast in Croatia), and Megleno-Romanian (spoken across a small area in Greece and Macedonia). Romanian itself is called Daco-Romanian or Northern Romanian when it is referred to as a member of this group. In this particular paper, the term “Romanian” will be used to refer to Daco-Romanian, unless otherwise indicated.

The languages in this Romanian subfamily belong to an ancient Romance continuum that stretched from the Iberian Peninsula to the Balkan Peninsula, but various migrations of peoples, notably the arrival of the Slavs in the Balkans, have left Romanian and its relatives isolated from their Romance “family.” The last surviving link between Romanians and other Eastern Romance variants (such as Sardinian and the various Italian dialects) was Dalmatian, spoken along the Croatian coast. The only surviving data from this language comes from a description of the language on the island of Krk, Croatia, published in 1906 and based on the speech of the son of the last fully native speaker.<sup>40</sup> However, the little data that is available does manage to place Dalmatian firmly along the

---

<sup>40</sup> Harris & Vincent, p.22

continuum as something of a link between Romanian and other Eastern Romance languages, though Romanian was actually isolated long before the death of the Dalmatian language.

#### **4.1 The Romance Language**

Romanian, due to this isolation, has a unique place in the Romance family. To the casual observer, as well as to the trained linguist, Romanian (and the various dialects throughout the Balkans) is unquestionably a Romance language in terms of both its diachronic development from Latin and synchronic typological criteria that are exhibited by all or most of the Romance languages. These include the morphological marking of noun gender, a rich synthetic system of verbal inflection (though in modern French the verbal system has become much more simplified due to phonological change and a trend towards analyticity), the apparent absence of a distinct modal verb morphosyntactic category, a complex system of clitic object pronouns, the use of definite and indefinite articles, and a system of pronominal verbs. One may also add a general SVO word order and a tendency to use external morphology, usually in the form of prepositions, to mark the nominal system (even in Romanian, the only nouns which are overtly inflected for case are feminine singular nouns – all other noun inflections rest on the enclitic definite article), and internal morphology in the form of synthetic paradigmatic morphemes to mark the verbal system.<sup>41</sup>

Of course, while morphological and syntactic typology are indispensable for strengthening the genetic classification of a language, it is the basic, *core* vocabulary which must determine its true heritage. The basic vocabulary of Romanian is very obviously composed of Romance stock: pronouns (*eu, tu, el, ea, noi, voi, ei/ele* and their

---

<sup>41</sup> Posner, pp. 36-37

various case forms), numerals (*unu, doi, trei, patru, cinci, șase, șapte, opt, nouă, zece*, though higher numbers are formed using these numerals combined in a Slavic fashion), question words (*ce* ‘what,’ *cine* ‘who,’ *unde* ‘where,’ *când* ‘when,’ and *cum* ‘how’), most prepositions (*spre* ‘above, on,’ *cătră* ‘against,’ *la* ‘at, to’ from ***illac-ad***, *înaintea*, ‘before, in front of,’ *între* ‘between,’ *de* ‘from, of,’ etc.), basic colors (*alb* ‘white,’ *negru* ‘black,’ *roșu* ‘red,’ *verde* ‘green,’ etc.), many other basic adjectives (*mulți*, ‘many,’ *alt* ‘other,’ *nou* ‘new,’ *vechi* ‘old,’ *bun* ‘good,’ *cald* ‘warm,’ *lung* ‘long,’ *larg* ‘wide’ etc.), names of common animals (*câine* ‘dog,’ *pește* ‘fish,’ *pasăre* ‘bird,’ *păduche* ‘louse,’ *șarpe* ‘snake,’ etc.), common plants and parts of plants (*arbore* ‘tree,’ *floare* ‘flower,’ *iarbă* ‘grass,’ *rădăcină* ‘root,’ etc.) body parts and products (*carne* ‘flesh,’ *os* ‘bone,’ *ău* ‘egg,’ *sânge* ‘blood,’ *cap* ‘head,’ *ureche* ‘ear,’ *ochi* ‘eye,’ *mână* ‘hand,’ *picioar* ‘foot,’ *piept* ‘breast,’ *păr* ‘hair,’ *ficat* ‘liver,’ *piele* ‘skin,’ *limbă* ‘tongue,’ etc.), common verbs (*a bea* ‘to drink,’ *a mânca* ‘to eat,’ *a merge* ‘to walk, go,’ *a se duce* ‘to go,’ *a vedea* ‘to see,’ *a auzi* ‘to hear,’ *a dormi* ‘to sleep,’ *a muri* ‘to die,’ *a ucide* ‘to kill,’ *a zbura* ‘to fly,’ *a bata* ‘to hit,’ *a fi* ‘to be,’ *a da* ‘to give,’ *a zice* ‘to say,’ etc.), common natural objects and phenomena (*soare* ‘sun,’ *lună* ‘moon,’ *stea* ‘star,’ *apă* ‘water,’ *ploaie* ‘rain,’ *mare* ‘sea,’ *pământ* ‘earth,’ *cer* ‘sky,’ *vânt* ‘wind,’ *foc* ‘fire,’ *munte* ‘mountain’) all come from Romance stock, namely Vulgar Latin.

All inflectional markers on nouns and verbs also come down quite transparently from Latin. The words *lup* ‘wolf’ and *casă* ‘house’ come from the Latin ***lupum*** and ***casam***, respectively (remember that modern “nominative” forms in Romance languages more often than not derive from the Latin accusative case form)<sup>42</sup>. The plurals *lupi* and *case* come from ***lupos*** and ***casas***, (the loss of the –s and its replacement with –i happened

---

<sup>42</sup> Posner pp. 119-120

early in Eastern Romance dialects). The genitive/dative of *casă* is *case*, from the Latin *casae*. The definite articles and the way they are marked for case in Romanian come from a disjunctive use of the Latin *ille*, ‘that,’ as they do in all other Romance languages. Romanian is unique in its enclisis of the article, due to its strong affinity within the Balkan Sprachbund. So Romanian has *lupul* from *lupum illum* and *casa* from *casam illum*. The case endings on the articles come from the various Latin case endings used for *ille* (ex. in *lupilor* ‘of/to the wolves,’ the article comes from *illorum*, and in *casei* ‘of/to the house’ it derives from *illi*. The loss of /l/ in the environment before a high front vowel is a common phonological feature of Romanian: *copil* ‘child’, *copii* ‘children’). Rosetti (1973) makes the rather dubious claim that the postposition of the article comes from the fact that adjectives follow nouns in Romanian, and that a Latin phrase like *homo ille bonus* (loosely: ‘man, that good one’) gives rise to the enclisis of the demonstrative (originally modifying the adjective) onto the end of the noun as an article – *omul bun*. But if this is the case, then why do all of the other Romance languages prepose a free article?<sup>43</sup> I should also mention that in adjectival phrases, the definite article will attach to the word that comes first, be it a noun or adjective (an adjective can sometimes precede the noun out of convention or to mark emphasis, as it can in other Romance languages), so ‘the good man’ can be expressed as *omul bun* or *bunul om*. This is not unlike other Romance languages except for the position of the article in relation to the word: Spanish *el buen hombre*, *el hombre bueno*.

---

<sup>43</sup> Rosetti (1973):

p. 39 : “La postposition de l’article, en roumain, est due à la place de l’adjectif après le nom : dans *homo ill bonus*, *ille* déterminait à l’origine l’adjectif ; ensuite, il a été rattaché au nom.

p. 57 : “En roumain et en albanais la postposition de l’article a été provoquée par la place de l’adjectif qui détermine le substantif “ He then makes the absurd claim: “*A la différence des autres langues romanes*, le roumaine place l’adjectif après le substantif “. This is of course, common to all Romance languages, not just to Romanian.

The morphology of the verbal system is also directly descended from Latin. The verb *a cînta* ‘to sing,’ comes from the Latin *cantare* (1<sup>st</sup> conjugation). Here is the

Romanian conjugation next to the Latin one:

1sg.	cînt	canto
2sg.	cînți	cantas
3sg.	cîntă	cantat
1pl.	cîntăm	cantamus
2pl.	cîntați	cantatis
3pl.	cîntă	cantant

And here are the Romanian and Latin conjugations for *a vedea* < *videre* ‘to see’ (2<sup>nd</sup> conjugation)

1sg.	văd	video
2sg.	vezi	vides
3sg.	vede	videt
1pl.	vedem	videmus
2pl.	vedeți	videtis
3pl.	văd	vident

Next, here are the Romanian and Latin conjugations of *a fi* / *esse* ‘to be’ (the Romanian word is derived from the Latin perfect stem *fu-*). This verb exhibits signs of other changes (such as the further extension of the form *sunt*), but the Latin origin is perfectly clear:

1sg.	sînt	sum
2sg.	ești	es
3sg.	e / este	est
1pl.	sîntem	sumus
2pl.	sînteți	estis
3pl.	sînt	sunt

The correspondence between Romanian forms of words and their Latin ancestors becomes even clearer when some of the sound changes that have occurred since the period of Vulgar Latin are explained. One of these changes is the development of two central vowels, *ă* (/ə/) and *â/î* (/i/), the difference between the two is simply orthographic,

the *â* being used to highlight Latin origins). The latter of the two is (or was historically) a very common sound in Slavic languages, and it usually considered to have entered the language through Slavic loanwords like *rîs* ‘lynx’ (from South Slavic *rysi*), and the sound was extended in words of Latin origin, as the phoneme /a/ (and sometimes /u/) became /ɨ/ in many instances where it preceded a nasal: *cîmp* ‘field’ < ***campum***, *lînă* ‘wool’ < ***lanam***. /ə/ was originally an unstressed allophone of /a/, and gained phonemic status under influence from other Balkan languages, as well as in the morphological alteration of indefinite/definite in pairs like *casă*, *casa*.

Another very prominent feature of the vowel system is the prevalence of the diphthongs *ea* and *oa*. These diphthongs arose when a stressed /e/ or /o/ assimilated the features of an /a/ (and consequently /ə/) in the following syllable. So, forms arise such as *seară* ‘evening’ < ***seram***, and *coadă* ‘tail, line’ < ***codam***. This also happened in a few words where the following vowel was /e/. For example, *floare* ‘flower’ comes from ***florem***. Initial /e/ in many cases has become iotacized into /ja/ or /je/, possibly under Slavic influence, so that *iarbă* ‘grass’ arose from ***herbam***, and *e* (/je/) arose from ***est***. Along with these vocalic changes, many of the unstressed vowels present in Latin were reduced or disappeared altogether, sometimes rendering the origins of a word quite obscure: *a zbura* ‘to fly’ comes from the Latin word ***exvolare***, dropping the first vowel altogether.

Some very early consonantal changes involved the labialization of labiovelars and of velars in certain contexts. One of these was the change of /kw/, /gw/ to /p/ and /b/ in most cases (almost all except for the question words and relative pronouns, which all retain the /k/: *când* ‘when’ < ***quando***, *ca* ‘how’ < ***quam***, *cât* ‘how much’ < ***quantus***, etc.)



so that *patru* ‘four’ results from *quattuor*, *apă* comes from *aqua*, *limbă* from *lingua*. This change also occurred in the obstruent clusters /ks/ and /kt/: *toapsec* ‘arrow poison’ < *toxicum*, *coapsă* ‘thigh’ < *coxam*; *drept* ‘straight’ < *directus*, *opt* ‘eight’ < *octo*, *noapte* ‘night’ < *noctem*. An intervocalic /l/ has undergone rhotacization in many cases: *cer* ‘sky’ < *caelum*, *moară* ‘millstone’ < *mola*. Finally, a Latin /v/ often resulted in Romanian /b/, so that Romanian has *corb* ‘raven’ from *corvus* and *bătrîn* ‘old (person)’ from *veteranum*.

Palatalization has played a major role in the development of the sound system of Romanian. Early on, /k/ and /g/ developed into /č/ and /j/ before front vowels, as they did in Italian (*cer* is pronounced /čer/, *merge* is pronounced /merje/). In many cases, alveolar stops broke into affricates when followed by a front vowel, so that *preț* ‘price’ comes from *pretium*, *țară* ‘country’ from *terram*, *zi* ‘day’ from *diem*, and *zece* ‘ten’ from *decem*. Sometimes /d/ could also produce /ž/ when followed by a glided high front vowel: *ajunge* ‘arrives’ from *adiungit*.<sup>44</sup> The same process occurred with the original Latin /s/, so that *șapte* ‘seven’ < *septem*, *și* ‘and’ < *sīc*, and *caș* < *caseum*. This also happens in cases in which the /s/ is separated from the high front vowel by an intervening stop: *trist* ‘sad’ < *tristem*, but *triști* < *tristes* (the phonological process here is somewhat complicated, but probably follows this sequence: modern *triști* /trištʲ/ < \*/trišti/ < \*/tristi/ < \*/tristei/ < *tristes* /tristes/). The phonemic status as well as the distribution of these sounds was augmented greatly by the influx of loanwords from Slavic languages, where these sounds are extremely common. The phoneme /l/ underwent a palatization before front vowels, and this resulted in its transformation into /j/ or its complete disappearance. Thus, we have *iepure* ‘hare’ < *leporem*, *ierta* ‘forgive’ < *libertare*, *fiu* ‘son’ < *filium*, and

<sup>44</sup> Note that the Romanian letter *j* is pronounced /ž/.

*muiere* ‘woman’ < ***mulierem***. This change also affected Slavic loanwords: *a iubi* ‘to love’ < Slavic *ljubiti*, and *nevoie* ‘need’ < *nevolja*.

In the modern Romanian language, palatalization is a phonemic process. The grapheme *i* at the end of a word is not actually pronounced, but indicates palatalization of the preceding consonant. So, for example, the word for ‘wolves,’ *lupi* is not pronounced \*/*lupi*/, but /*lup<sup>y</sup>*/, in opposition to the singular, *lup* /*lup*/. This process fully affects the velars, so that they break into alveopalatal affricates: *fac* /*fak*/ ‘I do,’ *faci* /*fač*/ ‘you do’; *merg* /*merg*/ ‘I go,’ *mergi* /*merj*/ ‘you go’ (there are a few exceptions, such as the word for ‘eye’ *ochi*, which always has a plural form, and is pronounced /*ok<sup>y</sup>*/). Most of the dental sounds are affected with more than just mere palatalization. The /*s*/ is typically mutated into /*š*/: *frumos* /*fru`mos*/ ‘beautiful (sg.),’ *frumoși* /*fru`moš<sup>y</sup>*/ ‘beautiful (pl.).’ A final /*t*/ will often change to /*ts*/ when *i* is added to the stem: *înot* /*î`not*/ ‘I swim,’ *înoți* /*î`nots<sup>y</sup>*/ ‘you swim.’ Its voiced counterpart, /*d*/, changes into /*z*/ rather than \*/*dz*/: *verde* /*verde*/ ‘green (sg.),’ *verzi* /*verz<sup>y</sup>*/ ‘green (pl.).’ The liquid /*l*/ is typically elided completely: *cal* /*kal*/ ‘horse,’ *cai* /*kai*/ ‘horses.’

The Romanian sound /*x*/ or /*h*/, written as *h* is not a continuation of the Latin ***h***, which disappeared very early in Vulgar Latin, but comes from the Slavic sound /*x*/. Such words as *duh* ‘spirit’ and *hrană* ‘food’ are loans from Slavic. In some cases, the /*h*/ changed to /*f*/, resulting *praf* ‘dust’ from Slavic *prahŭ*, and *vraf* ‘heap’ from Slavic *vrahŭ*. This process is not dissimilar to the changes involving the labialization of /*k*/ and /*g*/.<sup>45</sup>

---

<sup>45</sup> This entire section on the evolution of the sound system of Romanian is heavily indebted to the following sources for description of the processes as well as examples: Mallinson pp. 393-397, Du Nay pp. 51-52, 56-57, 62, 102, and Rosetti 24-26, 28-29, 69-71, 79-86.

## 4.2 The Differences

While the above section illustrates the conformity of the Romanian language to its genetic family, Romance, the language does of course exhibit many features which separate it and in many ways alienate it from its sister languages. In *The Romance Languages* (1996), Rebecca Posner seeks to describe the Romance languages synchronically using typological criteria. She cites Romanian as a “maverick language” on the fringe of the Romance continuum (French occupies the fringe at the other end).<sup>46</sup> Posner’s claims are based on the fact that at all levels of the language, Romanian exhibits features absent in the other Romance languages, either in retaining Latin forms that the others lost or in absorbing features of its neighboring languages. It is important to remember that the Romanian-speaking world existed outside of the sociocultural continuum that was the forebear to the Romance-speaking Western European nations. One must also consider that even at the time of the Roman occupation of the Balkans, Latin itself had already been fragmented into regional dialects with their own peculiarities. The isolation of Romanian effectively blocked all innovation from the center of the continuum, and this Eastern dialect of Vulgar Latin adopted new models for innovation, particularly Greek, Bulgarian, and other Slavic languages.

The vocabulary of Romanian is perhaps the most visible manifestation of these differences. Rosetti (1973) claims that of the 1000-1500 words that form the base of the Romanian vocabulary, 60% of the words have their origins in Latin, a low number when compared to the other Romance languages. In *Istoria limbii române*, edited by Rosetti, the authors give a pan-Romanic stock of 488 Latin words that are said to have comprised

---

<sup>46</sup> Posner, p. 40

the basic vocabulary of Vulgar Latin.<sup>47</sup> To get a clear picture of the differences in the Latin vocabulary of Romanian from the other Romance languages, one need only look at how these words were passed down into the modern Romance languages. Of the Pan-Romance word stock, 214 of the words are absent in Romanian. Most of these terms have to do with certain aspects of civilization, city life, religion, government, literacy, etc.<sup>48</sup> This suggests a very provincial lifestyle for the early Romanians, and when they did acquire words to fit these concepts, they came from other sources, such as Byzantine Greek and Old Slavonic. A comparison can be made here with English: most of the English vocabulary pertaining to these semantic spheres of civilization, religion, and government comes from French or Latin, not native English or even other Germanic sources.

There are also a significant number of fairly common words in Vulgar Latin that have survived in Romanian only. This is due to the fact that Romanian, already on the far end of the Romance continuum, did not participate in many of the semantic and lexical innovations that took place further to the West. Some of these words are: *iapă* ‘mare’ from *equam*, *mărgea* ‘bead’ from *margella*, *lingură* ‘spoon’ from *lingula*, *plăpînd* ‘weak’ from *palpabundus*, *împărat* ‘emperor’ from *imperator*, and *nici* ‘and not’ from *neque*.<sup>49</sup>

Romanian also has a disproportionately large number of “substrate” vocabulary items typically considered to be remnants of a pre-Roman Dacian / Illyrian population. These words constitute a good portion of the 10-15% of Romanian words that are of

---

<sup>47</sup> *ILR* p. 111

<sup>48</sup> Du Nay, p. 53

<sup>49</sup> Rosetti (1973), p. 43

unknown or unsure origin.<sup>50</sup> Many of these words are also present in Albanian, another Balkan language, and certain historians, including Du Nay (1996), have come to the conclusion that the numerous correspondences between the two languages derive from the same ancient language, the Albanians having kept theirs (though it has, of course, evolved) and the ancestors of the Romanians having become Romanized.<sup>51</sup> This sort of “substrate influence” is supported by the fact that many of the words belong to semantic categories that are not likely to have been borrowed by the Romanized population, but rather inherited from an ancestral language, such as body parts, kinship terms, housing, tools, etc. Also important is that there are many words in this category that have to do with animal husbandry and shepherding, reflecting the lifestyle of the pre-Roman inhabitants of the Balkan area. Du Nay asserts that it is likely that the Romanized population (which would have had to have been bilingual at some point) retained older terms where Latin lacked them.<sup>52</sup> Other authors are more skeptical of the substrate claims, such as Mallinson, who writes that “such cognates do not prove a common Thracian-Illyrian substratum any more than the fact that Albanian has many Latin loans cognate with Romanian means the two languages share a Latin heritage.”<sup>53</sup>

Regardless of the position to which one subscribes, some of the shared words are: Romanian *abure*, Albanian *avull* ‘steam’; *mînz*, *mës* ‘colt’; *scrum*, *shkrump* ‘ash’; *vatră*, *vatrë* ‘hearth’; *pîrâu*, *përrua* ‘brook’; *copil*, *kopil* ‘child (Rom.), bastard (Alb.)’; *ghiuj*, *gjysh* ‘old timer (Rom.), grandfather (Alb.)’, etc.<sup>54</sup>

---

<sup>50</sup> Du Nay p. 60

<sup>51</sup> Du Nay, p. 73

<sup>52</sup> Du Nay, p. 73

<sup>53</sup> Mallinson, p. 413

<sup>54</sup> examples come from Du Nay pp. 74-82, Rosetti (1973) pp. 58-63, and Mallinson, p. 413.

The last major group of Romanian vocabulary items which sets it apart from its Romance relatives is the staggering amount of borrowed Slavic vocabulary, which, along with Greek, filled much of the semantic void left by the lack of Latin terms in the aforementioned fields, and is much more substantial than the Germanic borrowings of the Western Romance languages.<sup>55</sup> Slavic loanwords, mostly from Bulgarian and other South Slavic languages (which at the time of the borrowings were almost identical anyway), constitute about 20% of the most basic 1000-1500 words,<sup>56</sup> and a very large number of the words that make up the learned vocabulary, though recent borrowing trends in the past 200 years or so have decreased this amount, as Latin, Italian, and French neologisms have replaced older Slavic borrowings. Slavic languages have had a tremendous impact on the Latin dialect that has come down to us as Romanian, starting in the sixth century and ending (or at least reducing to a trickle) in about the twelfth. They not only influenced the vocabulary, but also the phonology, derivational morphology, and in some cases even the syntax and inflectional morphology of Romanian.

Slavic words have entered the Romanian language in almost all semantic spheres, even in those which are typically most resistant to borrowing (kinship, body parts, etc. – this can actually be taken as a counterexample to Du Nay’s argument that cognates with Albanian are likely inherited and not borrowed). Some common Romanian words of South Slavic origin are: *pravilă* ‘law,’ *război* ‘war,’ *miset* ‘month,’ *prieten* ‘friend,’ *ceas* ‘hour,’ *bolnav* ‘sick,’ *a blagoslovi* ‘to bless,’ *hrană* ‘food,’ *plug* ‘plow,’ *bogat* ‘rich,’ *a citi* ‘to read,’ *a iubi* ‘to love,’ etc.<sup>57</sup> The Romanian word for ‘yes,’ *da*, is of Slavic origin.

---

<sup>55</sup> Mallinson, p. 413-414

<sup>56</sup> Rosetti (1973), p. 42

<sup>57</sup> Du Nay, pp. 104-105

In addition to numerous vocabulary items, Romanian has borrowed a number of Slavic prefixes and suffixes, many of which are quite productive and combine with Latin elements: *ne-* ‘not’: *nebun* ‘bad (not good),’ *nemulțumit* ‘unsatisfied’; *răs/z-* ‘[repeated action / detachment]’: *a răzbate* ‘to get through, pull through,’ *a răzgândi* ‘to change one’s mind’ (this prefix is etymologically related to the Latin *re-*); *-ean* ‘[forms ethnic adjectives and nouns]’: *sătean* ‘villager’ from *sat* ‘village’; *-enie* ‘[forms nouns denoting an action]’: *afuresenie* ‘excommunication’ from *afurisit* ‘accursed’; *-ic, -ice* ‘[diminutive]’: *pătic* ‘little bed’ from *pat* ‘bed,’ *gaurice* ‘little hole’ from *gaură* ‘hole,’ etc.

Romanian also differs from its Romance sister languages in ways which imply a deeper disconnection than simple vocabulary differences. In the areas of morphology and syntax, Romanian exhibits many features which distance it from its Western relatives, and by the same token, strengthen its ties to the Slavic and other Balkan languages with which it came into contact during the centuries of its formation. The most obvious manifestation of this is in nominal morphology. Other Romance languages typically have only one case in their nominal systems (perhaps two or three in their pronominal systems, however), but Romanian has at least two, as well as a third, which is more or less incomplete. The two cases are the Nominative/Accusative and the Genitive/Dative (the merger of the Genitive and Dative cases is a Balkan phenomenon, and is discussed above.) Romanian has, on the surface, preserved more of the Latin declension system than have any of its other descendants. That being said, it is most likely due to outside influence, rather than to any conservative nature within Romanian itself, that these cases were retained. Mallinson (1988) states that credit for the retention of case marking is in

part owed to the postposition of the definite article, based on the Latin determiner *ille*.<sup>58</sup>

If this is indeed the case, then we have one Balkan feature (enclitic definite article) reinforcing and supporting another (existence of a merged dative/genitive case). The following chart illustrates the basic breakdown of the Romanian case system in its three genders, two numbers, and two states of definiteness (the words are *lup* ‘wolf’, *casă* ‘house,’ and *oraș* ‘town’):

**Table 1.** The Romanian case system<sup>59</sup>

		<b>masc.</b>		<b>fem.</b>		<b>neut.</b>	
		<b>sing.</b>	<b>plural</b>	<b>sing.</b>	<b>plural</b>	<b>sing.</b>	<b>plural</b>
<b>Nom/Acc</b>	<b>-def</b>	<i>lup</i>	<i>lupi</i>	<i>casă</i>	<i>case</i>	<i>oraș</i>	<i>orașe</i>
	<b>+def</b>	<i>lupul</i>	<i>lupii</i>	<i>casa</i>	<i>casele</i>	<i>orașul</i>	<i>orașele</i>
<b>Gen/Dat</b>	<b>-def</b>	<i>lup</i>	<i>lupi</i>	<i>case</i>	<i>case</i>	<i>oraș</i>	<i>orașe</i>
	<b>+def</b>	<i>lupului</i>	<i>lupilor</i>	<i>casei</i>	<i>caselor</i>	<i>orașului</i>	<i>orașele</i>

It is obvious in the chart that only feminine singular nouns that are not definite actually get marked for case. However, the indefinite article, *un*, will be marked for case as needed, so that ‘of a wolf’ is expressed as *unui lup*. The neuter gender also exists in Romanian, but these nouns generally decline identically to masculine and feminine nouns, taking masculine forms in the singular and feminine forms in the plural. This is the default gender into which many new loanwords have been assigned for centuries, swelling the numbers of neuter nouns. As for proper names, one may use pronouns as markers of possession: *cartea lui Gheorghe*, ‘George’s book,’ or literally ‘the book of him George.’ This is actually how the definite article developed in the first place:

*cavernă lupilor* ‘the wolves’ cave’ < *caverna lupi illorum* ‘the cave, wolves, of those.’

There exists one more case in modern Romanian, absent in all other Romance languages: the vocative, which is gradually falling out of usage, but still alive,

<sup>58</sup> Mallinson, p. 400

<sup>59</sup> idea for chart taken from Mallinson, p.398



nonetheless. Only some nouns high in animacy (mostly people and animals) even show vocative case forms. The plural of the vocative case is the same as the definite genitive/dative form, and the singular is, for masculine nouns, the definite singular with the addition of *-e*, so that ‘hey man!’ is *omule!* (*om* ‘man’ + *-ul* + *-e*). However, most masculine kinship terms use the same form as the nominative. Proper names are either the same as in the nominative or take an *-e* ending: *Dane!* < ‘Dan!’ Feminine singular nouns sometimes change the final *-ă* to *-o* (*soră* ‘sister’ > *soro!*), betraying the true force behind this vocative case form: Slavic influence. Neither Classical nor Vulgar Latin had a vocative case form for feminine nouns, and the *-o* is clearly borrowed from Slavic languages (Polish, Czech: *Ana* > *Ano!*). The retention of the vocative case form in the masculine was most likely motivated by Slavic influence as well.<sup>60</sup> (Interestingly enough, it seems that the Romanian vocative *-le* form has been borrowed into Bulgarian and Macedonian: *bozhele!* *sestrole!* cf. Czech *bože!*, *sestro!*<sup>61</sup>).

In terms of verbal morphology, Romanian verbs typically have 5-6 forms (sometimes the 3<sup>rd</sup> person plural form is identical to the 1<sup>st</sup> person singular or the 3<sup>rd</sup> person singular), the norm for Romance languages (although French verbs only have 3 forms in the spoken language), and is, as is to be expected, a “pro-drop” language. The past tense is formed analytically, using a conjugated auxiliary ‘to have’ + past participle. So, while ‘I walk’ is *merg*, ‘I walked’ is *am mers*. However, oddly enough, Romanian has retained from Latin a synthetic pluperfect: ‘I had sung’ is *cîntasem*, deriving from the Latin pluperfect subjunctive *cantavissem*,<sup>62</sup> (this is odd since this tense is almost always formed analytically, even in the most synthetic Romance languages, Spanish and

---

<sup>60</sup> Mallinson, p. 400

<sup>61</sup> Du Nay, p. 109

<sup>62</sup> Mallinson, p. 407

Portuguese). The imperfect tense is also formed synthetically: *cîntam*, ‘I was singing,’ The future, in the literary language, is formed analytically, using a conjugated auxiliary ‘to want’ + infinitive. ‘I will go’ is *vreau a merge*. However, in the spoken language, the infinitive is all but a thing of the past, and spoken Romanian forms the future based on the typical Balkan paradigm (calqued from Greek): *O să merg*, where *o* is an invariant fossilized form of *vrea* ‘to want,’ *să* is a subjunctive particle, and *merg* is in the present subjunctive.

The infinitive is absent in many places where one would expect to find it in other Romance languages, replaced again by the subjunctive: ‘He must go’ is *trebuie să merge*, with the verb ‘to go’ in the subjunctive. One can compare this to, say, the Spanish *debe de ir* (but also *es necesario que vaya* ‘it’s necessary that he go,’ which has a different semantic connotation, one of ‘I/we desire that he leaves.’), the French *il doit aller*, and the Italian *deve andare*. Romanian can also use the subjunctive in sentences where one verb leads to another, like ‘he goes to see the man,’ rendered in Romanian as *merge să-l vadă pe omul* (lit. ‘he goes that-him he sees the man). This is clearly a Balkan feature, analogous to other Balkan languages (see above), and not found in the wider world of Romance languages (for example, the Spanish version of this sentence is *voy para ver al hombre*, which uses the infinitive). One place where the infinitive form does get used is in the conditional, where a conditional particle + infinitive construction is used: *aș merge* ‘I would go,’ *ar merge* ‘he would go.’

It is also important to note that the subjunctive is, at least in the present, almost an entirely analytic formation. The verb is only marked in the 3<sup>rd</sup> person singular form. All other persons and numbers are identical to the indicative. This means that *să* is actually

the key to the subjunctive, not the verbal morphology, as is the case in other Romance languages. In this way it is more like other Balkan languages, such as Bulgarian and Macedonian, which use the particle *da* for this purpose, and Albanian, which uses *të*.

While most Romance languages have complicated systems involving cliticized object pronouns, Romanian by far has the most complicated system. Double-marking through the use of clitic pronouns is commonplace and usually necessary, and fits in with general trends within the Balkan Sprachbund. For instance, in Romanian, the sentence ‘I saw her’ is *am văzut-o* (compare to Spanish *la he visto*, French *je l’ai vu*) literally ‘I have seen her.’ If the direct object is directly stated, as in *am văzut-o pe Maria* (Spanish *he visto a María*, French *j’ai vu Marie*), the direct object clitic pronoun appears in the sentence (literally, ‘I have seen her Maria.’). This redundancy for accusative objects is *obligatory*, especially when the direct object is animate. In the sentence *merge să-l vădă pe omul*, we can once again see this obligatory redundancy (lit. ‘He goes that him he sees the man,’). Also, regardless of the animacy of the noun, relative clauses will use a resumptive pronoun if the antecedent is a direct object in the relative clause: *cartea, pe care am cumpărat-o* (‘the book, which I bought it’; compare to Spanish *el libro que he comprado*), *omul, pe care l-am văzut* (‘the man, whom him I have seen him’; compare to Spanish *el hombre, que he visto*).

Spanish, out of all the other Romance languages, does have clitic constructions similar to some of those in Romanian. For example, when a *dative* is overtly expressed using the preposition *a* (accusative animates also take the preposition *a*, but they are treated syntactically as accusatives, not datives), one should place an indirect object clitic before the verb: *Le di un golpe a la pared* ‘I threw a punch at the wall,’ so that the literal

translation is ‘I threw a punch to it at the wall,’ complete with a redundant clitic pronoun (when the indirect object is not overtly stated, the sentence is *le di un golpe* ‘I threw a punch at it,’ with no redundancy). In phrases that use dative verbs, such as ‘I like *x*’ (‘*x* pleases me’), one can emphasize the dative object by using the dative prepositional phrase with *a*: *Me gusta* vs. *A mí me gusta*: ‘I like it’ vs. ‘I like it,’ literally ‘(to me) it pleases me.’<sup>63</sup> However, for accusative constructions only certain dialects of Spanish feature constructions such as *Le he visto a Juan* (standard *He visto a Juan*), which mirrors the Romanian *L-am văzut pe Joan*. This type of construction is not allowed at all in French: \**Je l’ai vu Marie*, except in cases where the object is dislocated, so that the sentence is *Marie, je l’ai vu*.<sup>64</sup> The sentence ‘you did not tell him everything’ in the three languages will demonstrate that Spanish and Romanian to reprise the dative, whereas in French this is ungrammatical: *lui nu i-au spus totul*, *a el no le has dicho todo*, but \**a lui tu ne lui as pas dit tout*. With accusative objects, standard Spanish aligns with French rather than Romanian: ‘When did the Romans conquer the Dacians?’ *cînd i-au învins romanii pe daci*?, but \**¿cuándo los han vencido los romanos a los dacios*?, \**quand les Romains les ont-ils vaincus les Daces*?<sup>65</sup>

Finally, while reflexive verbs are commonplace in all Romance languages, they are significantly more common in Romanian. This is because many of these verbs have acquired a reflexive particle based on a Slavic model even though the original Latin verb may not have had such a specification. In some cases they’ve also borrowed Slavic reflexive verbs wholesale, complete with the reflexive pronoun. Some examples are *a se*

---

<sup>63</sup> Posner, p. 168

<sup>64</sup> Posner, p. 168

<sup>65</sup> Gauger p. 12

*ruga* ‘to beseech, to pray’ < **rogare** (modeled from *moliti se* in Old Slavonic), *a se jura* ‘to swear’ < **iurare** (*kleti se*), *a se teme* ‘to fear’ < **temere** (*bojati se*), etc.<sup>66</sup>

Throughout this section I have intentionally avoided any talk of the actual history of Romanian, as far as its origins are concerned. In this section I have mainly been concerned with the linguistic classification of Romanian as a Romance language through shared inherited vocabulary and features, with allusions back to its classification as a Balkan language. The reason I have ignored history is not because it is not important, but because it does not fit here. The history of the Romanian language, unlike that of Italian, French, and Spanish, is by and large a mystery. There was literally a gap between the 3<sup>rd</sup> and 16<sup>th</sup> centuries from which we have no actual written records of Romanian. Given that Romanian exists in a part of the world which is charged with nationalism, ethnic tension, and until recently migration and conquest, the lack of strong evidence surrounding the language’s development turns any historical speculation into a political debate. Lines have been drawn, and one cannot make any claim about the history of the language during this period without making a rather strong political statement. The following section will explore the history of the Romanian language, the various theories of its development, and the political ramifications which arises from such a seemingly mundane task.

---

<sup>66</sup> Du Nay, p. 103

## 5 THE HISTORY OF ROMANIAN

The Romanian language is unique among Romance languages in that there is no continuous record of its development from Vulgar Latin. Whereas one may be able to find passages (whole documents or simple words and sentences) written in “the vernacular” throughout the history of Spanish or French for instance, the Romanian language effectively disappeared from history following the Roman period and did not reappear until 1521, in a letter written in the Cyrillic script by a local man to the mayor of Braşov. What happened during this time is a point of contention among historians, politicians, and ethnic groups. While we know that Romanian is a Romance language and consequently must be descended from Latin, the way in which it arrived in its more or less modern form in the 16<sup>th</sup> century and the path it took (quite literally, how its speakers migrated or did not migrate) to get there are hotly debated issues among historians and linguists alike, a debate that has been fueled more often than not by political and nationalistic agendas. Since the written record is so scarce in this case, linguistics has proven to be a valuable tool in attempting to reconstruct the past, as, according to Romanian linguist André Du Nay, “language is unconsciously transmitted historical evidence as is human remains, records, customs, etc.”<sup>67</sup>

Modern-day Romania occupies the extreme northern part of the Balkan Peninsula, separated from Bulgaria and Serbia by the Danube River, which plays an important role in the history of the Roman Empire and the history of the Romanian people. In ancient, pre-Roman times, the Balkan Peninsula was largely occupied by a loose grouping of Indo-European peoples. In the eastern part were the Thracians, and to the west lived the Illyrians. By the end of the 6<sup>th</sup> century B.C., these Thracians were called *Getae* by Greek

---

<sup>67</sup> Du Nay, p. 3



**Figure 2.** Modern Romania, with major regions, cities and geographical features (from *Wikipedia* with modifications)

chroniclers, and the northern group of them were later referred to by the Romans as *Dacii*, whence comes the English word “Dacians.”<sup>68</sup> By the 1<sup>st</sup> Century B.C., this area was a wealthy kingdom within the Roman area of interest. After more than a century of fighting and alliances with various Dacian kings, Rome finally entered Dacia in 101 A.D. under the orders of Emperor Trajan. By 106 Dacia was fully under Roman control, and would become an important source of grain and other resources, such as gold, lead, and salt. Dacia during the Roman occupation was known as *Dacia Traiana*, and it included much of modern-day Romania, including Transylvania and the Eastern regions of the

<sup>68</sup> Georgescu, p. 3

Banat and Oltenia (Interestingly, Roman Dacia did not include Wallachia or Moldavia, the two major regions which would eventually be the cradle of Romanian civilization).



**Figure 3.** Map of Roman Dacia (shaded and outlined in grey) with the borders of modern Romania outlines in black. (from Wikipedia with modifications)

Dacia was a rather prosperous province, but was under constant attack by Dacians outside Roman rule and their allies, the Sarmatians, as well as by rebellious Dacians living within the Roman borders. Soon, the Visigoths also entered onto the scene, and towards the end of the third century, the Roman army and administration gradually began to pull out of Dacia Traiana, moving back across the Danube to a more easily defensible position among the heavily Romanized regions to the south, which had already been under Roman rule for hundreds of years. By 275, the Romans had completely left behind Dacia Traiana, and with it most of its inhabitants, primarily farmers and shepherds. They maintained certain bridgeheads and outposts on the northern bank of the river, but for all intents and purposes, Dacia Traiana was abandoned by the Romans after about 170 years of provincial rule.



After this point in history, little definitive information exists about the region until the Middle Ages. During the period of time known as the *Völkwanderung*, in which peoples moved freely about Eastern Europe as the Roman Empire declined, Dacia was occupied by various waves of Visigoths, Huns, Avars, Cumans, other “barbarian” groups, and most importantly, Slavs (in the 6<sup>th</sup> century) and Magyars. The areas south of the Danube were similarly overrun in the early 7<sup>th</sup> century, mainly by Slavs, forever making the Balkans a largely Slavic domain.<sup>69</sup> Today there are still pockets of Latin speakers – they are the modern Aroumanians and Megleno-Romanians. These Romance speaking peoples south of the Danube are known throughout history as the Vlachs, a Germanic term which shares a common source with our words “Welsh” and “Walloon.” The Vlachs are part of the cast of ethnic groups that participated in recorded events on the Balkan Peninsula during the Middle Ages. South of the Danube, their existence is hinted at in various historical documents from about the 10<sup>th</sup> century onward (they are recorded in Northern Greece in 976, and “throughout Bulgaria” in the 11<sup>th</sup> century)<sup>70</sup>, but north of the Danube no record exists. No actual record exists of any group above the Danube until hundreds of years later. The lack of historical evidence from this period has allowed for the debate to rage beyond the confines of academic history and into the realm of politics and nationalism. There are two major theories concerning the origins of the Romanian language. One is the theory of continuity, which claims that the Romanian language and the Romanian people are direct descendants of the Roman occupation of Dacia Traiana, which would have resulted in a thorough Romanization able to withstand various invasions of alien peoples. The other theory is that of migration from the south, which

---

<sup>69</sup> Georgescu, p. 12

<sup>70</sup> Illyés, p. 251

assumes that a very weakly Romanized Dacia Traiana was consumed by its invaders, and that Latin-speaking peoples from the South (who could have also included the Romans who left Dacia) eventually moved across the Danube and that these people and their language were the basis for modern Romanian. I will not be vague in my evaluation of these theories: for various reasons that will be explained I find the second theory to be the more tenable one. The theory of continuity is part of a national mythology, a function of the Romanian nationalistic consciousness. This is why it is important in the context of this thesis: language is treated as a means to a nationalistic and political end; that end being the idea that the Romanians are the only indigenous people in the territory that is now present-day Romania.<sup>71</sup>

### **5.1 The History of the Two Theories and the Political Stakes**

Prior to the period of nationalism which swept Europe in the 18<sup>th</sup> century, very little regard was given to the question of Romanian origins at all. Then, while Transylvania was under Austro-Hungarian rule, members of an influential group of Romanian intellectuals living in that region, later known as the “Transylvanian School,” began to assert that the Romanians living within the empire deserved more autonomy in Transylvania because of their ethnic rights to Transylvania as the descendants of the ancient Daco-Romans. The region, with a mostly Romanian population, had long been under the control of Hungarians and Austrians, the Romanians themselves treated as backwards and provincial by the ruling government. This was one of the earliest incarnations of the theory of continuity, and reflected the views of Wallachian and Moldavian chroniclers from the previous century. In the 19<sup>th</sup> century, the German philologist Rupert Roessler first proposed the idea of a migration from the south, by

---

<sup>71</sup> Du Nay, p. 7

which the Romanians did not reach Transylvania until the 13<sup>th</sup> century. This theory has been upheld in the past by Hungarian scholars whose political motivations in Transylvania were no less transparent than those of the Romanians.<sup>72</sup> These theories lacked much objective scientific evidence until twentieth-century historians and linguists began to search for concrete support to their claims. Linguists were among the first Romanian intellectuals to question the theory of continuity, on the basis of the affinity Romanian shares with other Balkan languages to the south. The Romanian linguist Alexandru Philippide concluded in 1923 that the early Romanian language developed south of the Danube, based on Balkan linguistic evidence.<sup>73</sup> This view was shared by others, including Hungarians, Germans, and French.

However, throughout the 20<sup>th</sup> century the Theory of Continuity became axiomatic, and it was the official position of the Romanian Communist Party. Even today the theory of Daco-Roman continuity holds a strong position within the national consciousness of the Romanians. Many of the prominent Romanian historians and linguists throughout the latter half of the twentieth century support this theory. Among them are Constantin Daicoviciu, Alexandru Rosetti, Haralambie Mihăescu and Vlad Georgescu, who simply treat the theory of continuity as fact, and only mention counter-theories in passing. Even today, Romanians are sensitive to any claim that denies their “legitimacy” in the area. While either side may be debated, the fact that the historical record simply does not exist provides problems as well as loopholes for both theories. The theory of continuity is by and large nationalistically based, hinged on tenuous evidence and political agendas. However, as a result of the work of modern proponents of

---

<sup>72</sup> Castellan, p. 19

<sup>73</sup> Du Nay, p. 6

the theory, it has gained some degree of historical legitimacy. I will explore the theory of continuity and its criticisms, and while I hope to illustrate the importance of linguistic evidence in the question of Daco-Romanian continuity, I will leave it up to historians to solve or to debate the problem further. I shall make no claims that I or anyone else actually *knows* what happened; only that linguistic evidence strongly favors the view of a migration from the south.

## **5.2 The Theory of Continuity**

Before delving into the theory of continuity, I must emphasize again that the continuity discussed here involves the continuity of Latin-speaking peoples in Western Romania, that is, the part of Romanian which was the original Dacia Traiana. Most important is the question: were the Romanians living in the region of Transylvania at the time of the arrival of the Magyars in the 9<sup>th</sup> century? This question lurks at the very heart of any debate concerning which ethnic group has the rightful historical claim to Transylvania. And the truth is, even if Romanians were living in that region at the time, this does not necessarily mean that they are the same people who lived under Roman rule and spoke a Latin dialect, as they still could have migrated from the south during the previous 500 years. I should also mention that by the 11<sup>th</sup> century there is no doubt about the presence of Romance speakers north of the Danube, albeit in the regions which would eventually become Wallachia and Moldavia, which were outside of Dacia Traiana and at that point in the Bulgarian sphere of influence.

The first component of the theory of continuity that must be accepted before any others can even be considered is the idea that Dacia Traiana was a completely Romanized province, as was Gallia (France) or Hispania (Spain). Even though the Roman

administration and army deserted Dacia in the 3<sup>rd</sup> century, it is possible that the local population had become Latinized, persisting in the use of the language long after Rome itself was no longer the center of power. Dacia's population must have been bolstered by the arrival of Romans from various parts of the Empire, and it only makes sense that Latin would have become the language used between the various peoples who had no other common language. Georgescu (1984) writes: "The colonizing population was clearly heterogeneous, but whatever their origins, the colonists represented imperial culture and civilization and brought with them that most powerful Romanizing instrument, the Latin language."<sup>74</sup> By this argument, by the time the Roman army and administration left the colony in the 3<sup>rd</sup> century, they had left behind much of the Roman culture as well as the Latin language. Similarly, archaeological evidence of Roman style architecture and artifacts from before and after the Romans had left the colony is believed to indicate a strong Roman presence. But such evidence does not necessarily mean anything other than that the Romans were there and exerted a strong influence, which we know already from Roman sources.

However valid these claims seem to be, they only prove that there is a possibility that the population was thoroughly Romanized, not that it was.<sup>75</sup> Also, when the Roman army and administration left in the 3<sup>rd</sup> century, so did the driving force behind Romanization. For Romanization to continue, the common people left behind would have had to have been almost completely assimilated into Roman culture, to the point where the Roman element would have been strong enough to assimilate foreign invaders. This happened, for example, in the 9<sup>th</sup> century when Vikings began to settle in the northern

---

<sup>74</sup> Georgescu, p. 6

<sup>75</sup> Du Nay, p. 206, in response to arguments put forth by Daicovicu that were similar to the above arguments of Georgescu

part of France, resulting in Romance-speaking Normandy. However, the Romans had already been in that particular region for roughly 500 years by the time of the Vikings' arrival. There is also no indication that the Dacian language ever stopped being spoken during this period, and one cannot expect that the constant revolts by Dacians and invasions by free Dacians were conducted by a people who used the Latin language as extensively as subjugated peoples in Hispania and Gallia. Lastly, one need only consider the fate of Roman Britain to deduce that a strong Roman occupation and transference of material culture do not necessarily mean that a group of people has been thoroughly Romanized. Du Nay points out that "in spite of 365 years of Roman domination, abundant material remains of Roman style and customs, and even the preservation of dozens of Latin placenames, no Latin-speaking population survived in England."<sup>76</sup> There is no reason not to assume that a similar situation happened in Dacia, especially considering its much shorter occupation and its unsubmissive people, combined with a much heavier migration of non-Roman peoples through the area for a much longer time.

When the Slavs invaded the territory of modern Romania in the 6<sup>th</sup> century and crossed the Danube in the 7<sup>th</sup> century, they spread out all over the Balkans. This is why today most of the peoples who still live here are speakers of Slavic languages. According to the theory of continuity, these Slavs were assimilated into the Romance population north of the Danube, and to the south they assimilated the Romance population. Within this theory the Aroumanians and the Megleno-Romanians are remnants of the population along the Danube who wandered south.<sup>77</sup> This "assimilation" is regarded as evidence of the demographic and cultural superiority of the Daco-Roman population, which was

---

<sup>76</sup> Du Nay, p. 209

<sup>77</sup> Pop (1999), p. 31

supposedly still strong at the time.<sup>78</sup> Of course, the “evidence” of the assimilation itself is merely the fact that Romanian, not Slavic, is spoken north of the Danube. Such a claim can simply not stand as historical evidence of anything, since it does not prove whatsoever that Slavs were assimilated *at that time*. If the Slavs were assimilated, it could have also happened when the Vlachs from the south moved into the lands north of the river.

One can only ask why they were assimilated into Dacia, which was a briefly occupied Roman province already overrun by various ethnic groups, yet were able to assimilate the entire Balkan peninsula south of the Danube (excluding Greece and Albania), which was still a strongly Latin-speaking part of the Eastern Empire and had been so since before Dacia was ever even occupied. This question has simply not been addressed by most of the major proponents of the theory. To their credit, however, there is also no necessary proof that if Romance speakers survived above the Danube, they were assimilated into invading groups. Conquerors do not always bring their language with them, but often rather assume the language of the native population, which is what happened in Normandy in the above example (and also in Spain and Italy, after Goths had conquered both of these regions). While we know that most of the population south of the Danube were in fact Slavicized (they certainly all speak Slavic languages today), it is possible, though I consider it unlikely, that the population north of the Danube absorbed the Slavs into a Romance culture and language group. Those who support a theory of migration from the south would claim that if the Romance population even existed at that point, it would have been completely overtaken by Slavs and other ethnic groups passing through. It would not be until remnants of the Roman population south of

---

<sup>78</sup> Oțetea, p. 191

the Danube (Vlachs) living within the Bulgarian empire traveled across the Danube that a Romance language would reenter this region, at the locations of present day Wallachia and Moldavia, and only then travel into Transylvania, which may or may not have been occupied by Hungarians yet.

Another claim of the theory of continuity is that the Romance dialect that would eventually become Romanian would have been preserved in the Carpathian Mountains and in the Transylvanian plateau, spreading into the regions of Wallachia and Moldavia in later centuries. Certainly this is not unheard of throughout Europe: rough geography is partly responsible for the continued existence of the Basques, the Welsh, and the Scots, to name a few. One of the bases for this idea is the 12<sup>th</sup> century chronicle *Gesta Hungarorum*, the chronicles of the exploits of the Magyar people, written by an anonymous scribe of the Hungarian king Bela III. In it, the writer describes the arrival of the Magyars in Transylvania (in the 9<sup>th</sup> century), and their encounters with local peoples, who included Vlachs, Bulgarians, and Slavs.<sup>79</sup> The Hungarian chronicler is apparently claiming that the Vlachs were living in this area when the Hungarians arrived. But all this would mean is that Vlachs were there at the time, and if we were to take the writer's claims into consideration, so were Bulgarians and other Slavs. Nothing in the record is there to indicate that the Vlachs mentioned are the same Romance-speakers left behind in Roman Dacia. Also, the credibility of this anonymous source is questionable. How much could the writer, compiling his work in the 12<sup>th</sup> century, know about the native populations of Transylvania 300 years before? It is much more likely that he was projecting the ethnic situation as he knew it to be in his own time. And while *Gesta Hungarorum* is considered to be a reliable source by some, it must be seen for what it

---

<sup>79</sup> Georgescu, p. 14



really is: an embellished “history” of the foundation of the Hungarian kingdom written according to the fashion of the time, based on a combination of current events and oral histories.<sup>80</sup> It can be considered an historical document only with an enormous amount of suspicion.

### 5.3 The Theory of Migration from the South

While the theory of continuity has support among many Romanian historians, there is still no undeniable evidence that the modern Romanians are the descendants of the Daco-Romans of ancient times. The historical record is vacant, so there is no evidence there. Archaeology does not prove much since it is unlikely that all vestiges of Roman life and technology would have disappeared after the Romans left anyway. The very few Latin inscriptions from the period after the Roman withdrawal are easily explained away. For instance, one of these is a Christian votive offering on a bronze tablet, with the inscription *ego Zenovius votum posui* (“I, Zenovius, made a votive offering”)<sup>81</sup>. However, such Latin inscriptions on objects have been found all over Europe from this time period, even in parts where the Romans themselves had no presence, such as in Scandinavia and parts of modern Germany.<sup>82</sup> All this indicates is the presence of someone who knew Latin, perhaps a missionary, or even someone who took an object from some other region of the Empire. It does not represent anything about the language of the general populace.<sup>83</sup> And finally, even if Romanians were north of the Danube at the time of the arrival of the Hungarians in Transylvania, this still does not prove that they are a continuation of the population living under the Romans.

---

<sup>80</sup> Du Nay, p. 219

<sup>81</sup> Georgescu, p. 11

<sup>82</sup> Du Nay, p. 221

<sup>83</sup> Illyés, p. 83

The modern theory of migration from the south, however, is primarily based on linguistic evidence. While physical and historical evidence for the origins of the Romanians is unclear or nonexistent, linguistic evidence is abundant in the speech of the millions of people living in the Balkans today who speak a Romance idiom, be it Romanian, Aroumanian, or Megleno-Romanian. While other theories of migrations from other regions have been proposed by non-Romanian scholars and officials for political reasons (usually to support Hungarian rights to Transylvania or Soviet / Russian rights to Moldova), the modern theory is based on the same linguistic evidence on which Alexandru Philippide based his claims that Romanian must have developed south of the Danube and was brought into modern Romanian at a later date. This evidence is all given in the framework that no one really knows what happened historically, and the most concrete evidence that exists today is in the language itself. The theory is still met with open hostility today, but this is more an emotional response rather than a reasoned argument. Ioan-Aurel Pop (1996), for example, writes “the assumption that the Romanians emerged in a place other than the one they have inhabited until now (the Roman province of *Dacia*) can be made only in case of deliberate ignorance of historical sources and common sense analogies.”<sup>84</sup> But even as authors make such strong assertions, they themselves deny very strong linguistic evidence in favor of tenuous and entirely inconclusive archaeological findings.

To begin, one must consider the demographics of the Romanian language today. Romanian is spoken by over 23 million people in Romania, Moldova, and the Ukraine, in a contiguous area larger than all of its neighbors. For the past four centuries or so since the first actual appearance of Romanian in print (the first attestations of the language),

---

<sup>84</sup> Pop (1996), p. 29

Romanian as it is spoken throughout this area has exhibited very little dialectal difference whatsoever. The Romanian language, while it did undergo standardization, was never subject to the great amount of leveling or supercedence of a particular dialect as seen in other nations such as France and Spain. However, such leveling has never really been necessary for Romania, since ever since Romanian has appeared in writing, it has exhibited minimal dialectal variation. While there are some regional distinctions (mostly in accent), of course, there exists by no means the kind of variation one would expect in an area that had no central organization until the 19<sup>th</sup> century (as can be seen in the dialects of Italy and Germany). The country in which Romanian is spoken was never even politically unified until the 19<sup>th</sup> century. Therefore, to believe the theory of continuity, one would have to believe that the language of the Daco-Romans developed in a more or less uniform way across a vast area north and south of the Danube without common political boundaries (including places that were never Roman territories, such as Moldavia and Wallachia), without large urban centers for hundreds of years, without political forces driving linguistic unity, while all the while being constantly bombarded by outside influences and governed by foreign powers. Yet, this is what one *must* believe if one is to adhere to the theory, as many Romanian historians and linguists do.<sup>85</sup> The theory of migration from the south maintains that Romanian developed in a much smaller area south of the Danube, and then was brought across it hundreds of years after the Romans had abandoned Dacia. The location south of the Danube is based on correspondences between Romanian and other languages in the region, as well as on the absence of certain linguistic elements, particularly Gothic.

---

<sup>85</sup> for example: "It was thus that a Romanic people arose over extensive areas before the early 7<sup>th</sup> century under definite historical conditions" -- Oțetea, p. 131

The theory of migration from the south is also supported in the Latin elements of Romanian itself. Romanian exhibits certain innovative features which prove that its Latin speaking forebears were still in regular contact with the rest of the Romance world until the early years of the 7<sup>th</sup> century. These features, many of which are shared with Italian dialects and Dalmatian, are indicative of changes that took place in Late Latin after the 3<sup>rd</sup> century. Some of the features include the palatalization of /k/ and /g/ and the assibilation of /t/ and /d/ before /e/ and /i/ (Latin *vicinum* (/wikinum/), *terram* (/terram/) > *vecin* (/večin/), *țară* (/tsarə/) ‘neighbor,’ ‘land’), the perfect verb tense formed using a *habere* auxiliary, a wealth of vocabulary, including many Christian terms (which Continuity supporters claim to have entered the language north of the Danube), as well as particular semantic shifts, such as the use of *manduco* for ‘to eat’ instead of a variant of *edo* (Italian *mangiare*, Romanian *a mânca*).<sup>86</sup> Roman culture itself had been thriving in the areas south of the Danube, and continued to do so until the region was overrun by the Slavs around 602. Of course, these correspondences with later developments in other Romance languages are used by continuity theorists to “prove” that Latin speakers north of the Danube were in continuous contact with those to the south, but this view is unlikely, as there is little archaeological evidence of large-scale contact across the river. Besides, such innovations are explained much more convincingly by a theory which places the early Romanians in an area where Romance contact at this time can be taken for granted.

One of the major reasons why Romanian is believed by many to have developed south of the Danube is its place within the Balkan Sprachbund. Romanian, as mentioned above, is a full member of this linguistic union, along with Bulgarian, Macedonian and

---

<sup>86</sup> Illyés, pp.202-207

Albanian. The Balkan Sprachbund involves certain morphological and lexical features of untraceable origin, as well those heavily borrowed from local languages, particularly Greek. Romanian occupies a place in the inner circle of the Sprachbund; yet it is geographically located further away from the other languages than are Serbian and Greek, themselves only peripheral members of the linguistic union. In other words, Romanian is more of a Balkan language than Greek or Serbian; yet is geographically further removed, and is in fact separated by one of Europe's major rivers. While perhaps not impossible, it is highly unlikely that Romanian would have developed such a strong affinity with the other Balkan languages if it actually developed so far away from them. This is one of the major reasons why proponents of the theory of migration from the south place the Romanian homeland somewhere roughly corresponding to an area in Macedonia. Those who support the theory of continuity (such as Rosetti),<sup>87</sup> unable to deny the facts concerning the Balkan Sprachbund, typically write it off as evidence of a very large area of formation, including areas both north and south of the Danube. They may also consider such Balkanisms to be parallel results of a common substrate language group. But, many of the innovations are directly attributable to Greek, and the others are focused around a rather small epicenter. And whereas Bulgarian, Serbian, and even Albanian dialects all show diminishing frequency in Balkanisms the further they are removed from the epicenter of the Sprachbund, Daco-Romanian is uniform in its usage across the entire area in which it is spoken, with no one dialect containing significantly more or fewer Balkan traits than another.<sup>88</sup> The only logical conclusion, then, is that Daco-Romanian developed somewhere closer to the center of the symbiosis of the Balkan

---

<sup>87</sup> Illyés, p. 194

<sup>88</sup> Du Nay, p. 258

regional family, and only then moved out. This is the only way to explain the uniformity in regard to Balkan characteristics that one will encounter all the way from the south in Bucharest, Wallachia to the north in Chişinău, Moldova.

Even more striking is the affinity that Romanian shares with Albanian, whereby the two languages share lexical items and other characteristics (see above) not found in Bulgarian or in other Balkan languages. Today, throughout the entire Balkan Sprachbund, the two languages separated by the most geographical distance are Albanian and Romanian. However, these languages share a significant amount of vocabulary: of the roughly 209 words assumed to come from substratum sources in Romanian, 113 have counterparts in Albanian.<sup>89</sup> While proponents of the theory of continuity insist that the affinities between the two languages simply come from a common, pre-Romance substrate stock, is a modern linguist to believe that such politically and geographically fragmented groups of ancient peoples shared a more or less common and unitary language? Regular sound change indicates that at the time the Latin dialect that would eventually lead to Romanian acquired these words, their form was more or less identical to the form existing in Albanian at the time. So, for instance, the only sound changes that separate Romanian *abur* from Albanian *avull* “steam” are those which are known to have occurred during the evolution of Latin to Romanian. Many such words exhibit no difference in pronunciation, such as Romanian *copil* “child” and Albanian *kopil* “bastard”, or *ciucă*, *çukë*, both meaning “peak, summit”.<sup>90</sup> This either implies direct borrowing from Albanian or a common substrate with Albanian. Both of these possibilities would necessitate that Romanian had developed in the vicinity of Albanian,

---

<sup>89</sup> Illyés, p. 242

<sup>90</sup> Du Nay, p. 80

not hundreds of miles away and across a major river. To suggest that there was a common substrate extending from far north of the Danube all the way down the Illyrian coast would simply be theoretically unsound and more or less preposterous. While the languages spoken across this region were related, perhaps even closely related, it is highly unlikely that these words would have been so similar in phonological form by the time Romanian developed, especially considering that the languages had already been in place for perhaps more than a thousand years.

The Slavic element in Romanian accounts for roughly 20% of the most basic lexical items in the language.<sup>91</sup> This number is not to be taken lightly, as it is much more substantial than the Germanic adstrate in Romance languages further to the West.<sup>92</sup> Romanians, wherever they may have come from, have lived side by side with Slavs for centuries, and such a massive amount of borrowing is to be expected, regardless of whether the Romanians came from north or south of the Danube. However, one must remember that by the early Middle Ages, the more or less unitary proto-Slavic language had at least broken up into three major branches: South Slavic, West Slavic, and East Slavic. Each group possesses certain phonological, morphological, lexical, syntactic, and semantic traits which separates it from the others, due to regular sound change and other linguistic shifts. The Romanian language contains Slavic elements from two main periods, which we are able to identify because of what is known about historical sound change within the Slavic languages: the period of the 7<sup>th</sup>-8<sup>th</sup> centuries, when Slavs had first broken across the Danube and into the Balkan Peninsula proper, and the period during the 11<sup>th</sup> and 12<sup>th</sup> centuries. Most of the Slavic loans during the second period come

---

<sup>91</sup> Rosetti, p. 42

<sup>92</sup> Mallinson, p. 413

from what appears to be Middle Bulgarian,<sup>93</sup> as the Bulgarian nation was rising to prominence in the Balkans, and Vlachs (Romanians) first started to really appear as one of the major ethnic elements within the Bulgarian kingdom. Interestingly, modern Northern Romanian contains a much heavier Bulgarian element than does its relative to the South, Aroumanian and Megleno-Romanian. This is thought to be due to a much tighter symbiosis resulting from Romanians and Bulgarians living among one another south, and later north, of the Danube during the Middle Ages.

If the theory of Daco-Roman continuity is correct, then the early and most substantial Slavic influence should not have come from the South Slavic group, of which Bulgarian is a member, but from the Eastern (such as Ukrainian and Russian) and Western (Czech and Slovak) branches, as these would have been the Slavs settling and mingling among the areas of Transylvania in the west and Moldavia in the east. As it is, various toponyms from Transylvania do exhibit Western Slavic sound patterns, so one would expect Slavic loanwords coming from this era to also exhibit such features if they did indeed arise in this area. However, out of the Slavic wordstock in Romanian, lexical items of East and West Slavic character do not appear until the 12<sup>th</sup>-13<sup>th</sup> centuries.<sup>94</sup>

Finally, there is the issue of toponyms, place-names. None of the place-names in Romania north of the Danube today are of an inherited Latin origin, the 11 or 12 Latin urban centers having been abandoned soon after the Romans' departure.<sup>95</sup> One would certainly expect that place-names would persist if the Romania, particularly the Transylvania, of today is the descendant of Roman Dacia. While this in itself proves nothing, it is just one more anomalous point that must be explained away by the theory of

---

<sup>93</sup> Du Nay, p. 241

<sup>94</sup> Du Nay, p. 241

<sup>95</sup> Illyés, p. 197



Daco-Roman continuity. In *Istoria limbii române* (1968), Rosetti proposes the following explanation: “the fact that the majority of Rumanian placenames north of the Danube are Slavic is explained by the fact that the very numerous Slavic population translated the older names of villages, as Frumoasa into Dobra, Piatra into Kamenu and that the towns were founded by foreigners [he is referring here to towns founded after the exodus of the Roman administration]”<sup>96</sup>. Such a proposition, while possible, is highly dubious: that foreigners just renamed all of the villages in what was supposedly a majority Daco-Roman land, and the population just accepted it as such. In fact, it is much more likely that the Slavs would have simply adapted the names into their own phonological systems and used the old names, as they did in Serbia, where the names of many towns and villages betray their Romanian or Latin origins.<sup>97</sup> For a more familiar example closer to home, one need only consider the hundred of place-names of Roman origin in modern England that were preserved through Old English, so that for instance, we get the toponymic elements “-caster” and “-chester” from the Latin word for army camp, *castra*.<sup>98</sup> The Romans themselves often preserved the original Celtic names of places in England and France (some examples include Kent and Canterbury). In the United States and in Africa, there is an abundance of names of settlements, regions, and geographical features which were preserved from their original languages by European colonists, even though many of these languages have long become marginalized or even extinct. Another important point is that none of the names of the major rivers that run through Romania exhibit a sound pattern which identifies them as deriving from Roman names. All of the names of the rivers, whether or not they were known in Roman times, have sound

---

<sup>96</sup> Rosetti, ILR, p. 215; translation quoted in Du Nay, p. 192

<sup>97</sup> Du Nay, pp. 30-31

<sup>98</sup> Du Nay, p. 247

patterns which betray them as borrowings into Romanian from Slavic or Hungarian. The Romanian names reflect these forms, meaning that they had to have acquired these names from another source besides their own local variety of Latin.<sup>99</sup> When one takes all of this into consideration, there does not seem to be sufficient evidence to explain the utter absence of Latin toponyms north of the Danube in order to support the theory of Daco-Roman continuity.

#### **5.4 The Final Verdict: Which Theory Is More Solid?**

Both theories do present interesting supporting arguments. But I believe that in the end, the linguistic data in support of the second theory simply makes the first one untenable. The theory of Daco-Roman continuity suggests that a briefly occupied and unstable region on the far end of the empire was thoroughly Romanized, enough to withstand invasions of various groups, some of which settled in the area. It suggests that the Romanian language was formed throughout an area that was larger than present-day Romania, encompassing areas both north and south of the Danube which had been separated politically from the 3<sup>rd</sup> century until around the 12<sup>th</sup>. The theory allows for correspondences between the Balkan languages, particularly Albanian, to be explained by the presence of a vast substratal language of a more or less unitary character. Lastly, it assumes that this vast area was finally broken up as Slavs swept through the Balkans, assimilating the Romanians south of the Danube (the ones with a better-established Roman history), or driving them further south, and being assimilated themselves into the culture of the Romanians north of the Danube.

The theory of migration from the south suggests that Roman Dacia may or may not have been completely abandoned by Latin-speakers, but that whatever the case,

---

<sup>99</sup> Illyés, p. 317

modern Romanian is not the direct descendant of the language spoken in Roman Dacia. It asserts that in fact, the modern Northern Romanian language and the other Balkan Romance languages developed from Latin in a region south of the Danube, probably in Macedonia, where either through borrowing or common substrate, it acquired words with matching forms in Albanian, as well as other Balkan features, it being in the epicenter of the zone in which the Balkan Sprachbund was formed. Eventually, the speakers of the dialect that was to become Northern Romanian permeated the highly cosmopolitan Bulgarian kingdom (which controlled lands on both sides of the Danube), where another Slavic layer was added to the language, this time including many words involving city life and religion. From this Bulgarian kingdom, Romanians found their way north across the Danube, eventually disseminating into the regions of Wallachia, Moldavia, and Transylvania, and speaking a very early form of what would become Modern Romanian. Meanwhile, the speakers of the languages that would eventually become Aroumanian and Megleno-Romanian remained in the south and continued the traditional Vlach pastoralist lifestyle.<sup>100</sup> This also explains why there is a much more intense Slavic element in Northern Romanian than in the Southern dialects.

To the linguist, the theory of continuity involves too many leaps of faith to seem comfortable. It leaves entirely too much to be explained. On the other hand, the theory of migration from the south is hard to reconcile with the fact that Romanians are the largest and most widespread nationality in Southeastern Europe (except for Ukrainians). Both theories had origins in the political ideas and motivations of Romanian and non-Romanian leaders, drawn up in a time when nationalism ruled Europe, especially the Balkans. Such nationalism did not affect only the Romanian conception of history, but

---

<sup>100</sup> Illyés, p. 252

also, obviously, the Romanian conception of the language. In time, the Romanian language would become testament to the fact that nationalism and ethnic pride can become the foundation for language change itself.

## 6 STANDARDIZATION (AND MANIPULATION) IN THE WAKE OF ROMANIAN NATIONALISM

By the beginning of the 18<sup>th</sup> century, the Romanians had been long established as a people in the regions they now inhabit. However, the nation-state of Romania did not yet exist. Instead, one may think of the predecessor of modern Romania as three separate entities: Wallachia, Moldavia, and Transylvania. At this time, none of these principalities



**Figure 4.** “Romania” at the turn of the 19<sup>th</sup> century (from *Wikipedia*).

enjoyed political independence from the great powers of the region. The Romanians were, in effect, a people without a nation, with none of the pomp and grandeur of the neighboring empires. They were an ethnic group lost in the middle of the shifting power struggle that characterized the Balkans in that period, whose effects are still felt to this day. The Romanians of this period were a proud but denigrated people, mired in political poverty when compared to their overlords and neighbors. In the West, the Austrian

Hapsburg Empire (including the kingdom of Hungary) dominated Central and Eastern Europe; in the East, Russia exercised its might around the Black Sea; and in the South, the Ottoman Empire lorded over the former vassal states of the Byzantines. Transylvania at this time was ruled by the Hapsburgs, while Wallachia and Moldavia were principalities under the rule of the Phanariots, Byzantine Greeks answerable to the Ottoman Sultan. The Romanians were small players on the scene during this era, but Romanian intellectuals and activists in all three of the principalities were already beginning to stoke the fires of nationalism.

The rise to nationhood will unequivocally result in a standardization of the national language, notably in orthography and grammar. Romania is odd in that the local varieties within the nation exhibit only minor amounts of variation, so that standardization did not necessarily result in one dialect subsuming the rest. One may compare this with Spain, where the Castillian dialect eventually overtook Aragonese, Navarrese, Leonese and other local dialects / languages in the name of standardization. The standardization of Romanian, on the other hand, was guided by intellectuals with one major, overarching goal: to emphasize and in many ways augment the Latinity of the language – for if the Romanian people had the Latin language as their birthright, then surely the prestige of Romance culture and the political legitimacy of a long tradition also belonged to them.<sup>101</sup> In the eyes of the reformers, Romania's destiny was to reclaim its place within Romance and the former glory of the Roman Empire. Even today, Romanians are deeply proud of their Latin heritage, and statues of Romulus and Remus adorn cities and towns throughout Romania.

---

<sup>101</sup> “They were attracted by the notion that their descent from the Romans gave them a special right to the rich cultural heritage of the Romance-speaking peoples” – Close, p. 23

It is at this time in a language's history that its natural development becomes muddled by conscious change, when man and not circumstance becomes the driving force behind the linguistic canon. Like all other European languages, Romanian entered a period of transformation, to become a tool for political purposes and not just for day-to-day communication. And so, internal prescriptivism became commonplace in the 18<sup>th</sup> and 19<sup>th</sup> centuries, enough to change the language itself into the form in which it is now known today. The standardization of the language centered around two intellectual circles, one in Hapsburg-ruled Transylvania and the other in Ottoman-dominated Wallachia and Moldavia. These two will be dealt with separately.

## **6.1 Transylvania**

Transylvania is a region of mixed ethnic character, comprised of Romanians, Hungarians, Saxons, and Gypsies, among other groups. It had long been the dominion of Hungary, ever since it was incorporated into the Hungarian kingdom in the 11<sup>th</sup> century. Hungarians have historically considered Transylvania their own, and it maintains an important place within their historical outlook. This land was the land which enjoyed a golden age of practical independence as an Ottoman client and a stronghold of Protestantism against the Hapsburgs in the 16<sup>th</sup> and 17<sup>th</sup> centuries. Hungarians had fought hard and died to protect this region, and it was the site of the last stand in the Hungarian Revolution in 1849 when Hungarians fought against both the Austrians and the Russians for independence.<sup>102</sup> Romanians, on the other hand, had for centuries maintained an ethnic majority in the region. To Romanian nationals, Transylvania was their ancestral homeland of Dacia Traiana, the birthplace of their civilization. For these reasons, Transylvania has historically been a hotbed of political debate and polemic, as each side

---

<sup>102</sup> Hupchick and Cox, map 38.

accused the other of trying to usurp its claims to the area. The debate has continued well into modern times, as Transylvania changed hands between Romania and Hungary three times in the 20<sup>th</sup> century. Today, Transylvania remains securely within Romania.

Romanians in particular, a scattered people living within the borders of various European nations, bore the brunt of much denigration, especially in Transylvania. In part, the development of the national self-image of Romanians in Transylvania, as in greater Romania, is due to vehement opposition to the degrading attitudes of their more powerful neighbors and overlords. One must remember that this was a time where ethnic diversity and tolerance were frowned upon, and the Romanians, being the weakest of three major ethnic groups in the region, were seen as a “problem” by Hungarian and Austrian authorities alike.<sup>103</sup> It was this environment which would give rise to a small group of historians and linguists who sought to empower the Romanian people against outside domination through the development of a language and culture which was undeniably “superior” to those of the “barbarian” Hungarians and Austrians.

In 1791, Romanian Transylvanians issued the *Supplex Libellus Valachorum*, a document which demanded and justified rights for Romanians as an ethnic and a linguistic group, based on the antiquity and numerical superiority of the Romanian people. The contents of the *Supplex* were partly taken from the ideas of a Romanian bishop from the earlier half of the century, Inochentie Micu Clain, who had demanded equal rights for Romanians with other peoples living in Transylvania (namely Hungarians

---

<sup>103</sup> This would culminate in the strict Magyarization policies instituted later in the 19<sup>th</sup> century in an effort to create cultural homogeneity. As one official put it: “Inside the Magyar state there can be no nationalities, nor nationality rights. There is only one nationality: every Hungarian subject is Magyar”. Hungarian eventually became a mandatory subject of study in Transylvanian schools, and Romanians and others were encouraged to even go so far as to Magyarize their family names (Bodea and Căndeia, p. 57-59)



and Saxons), such as the right to free movement, access to education, and representation in the government. The bases for his claims were the antiquity, continuity, and therefore cultural superiority of the Romanian people.<sup>104</sup> The *Supplex* itself asked for the restoration of the ancient rights deserved by the Romanian people, and insisted that the time was right for such a restoration, as the Romanian population had reached one million out of the 1.7 million people living in Transylvania.

While many intellectuals contributed to the *Supplex*, there are three who stand out above all others: philologist Samuil Micu, and the historians Gheorghe Șincai and Petru Maior. These men had studied in Hungary, Vienna, and Rome, participating in many of the important cultural events and movements of late 18<sup>th</sup> century-Europe. Their education provided them with an extensive knowledge of European history, mastery of Latin philology and other Romance languages which gave them a new perspective on their own language, and with the overall philosophy of the Enlightenment, which was responsible for the ideas behind both the American and French Revolutions. Together, these men became known as the *Școala Ardeleană* – The Transylvanian School. Their work contributed greatly to not only the affirming of national rights of Transylvanian Romanians, but also provided a significant intellectual jump-start to the entire Romanian nationalist movement, which would result in the creation of the Romanian state in the 19<sup>th</sup> century (which, ironically, did not include Transylvania).<sup>105</sup>

The members of the Transylvanian School were some of the earliest and most dedicated proponents of the theory of Daco-Roman continuity, and they backed up their arguments with what passed as scientific, historical, and linguistic facts in the 18<sup>th</sup> century. While

---

<sup>104</sup> Bodea and Căndea, p. 36

<sup>105</sup> Bodea and Căndea, p. 37

many of their stances on the history of Romanians were rejected even by other Continuitists, such as the idea that Dacia was inhabited solely by Romans as the Dacians had all been killed off,<sup>106</sup> the spirit of their efforts was enough to bring intellectual inquiry and scholarship into a nationalistic debate, helping to form the Romanian conception of their own history and language, as well as to contribute to the standardization of the Romanian language. All of their linguistic arguments and innovations were based on the idea that the Romanians were the linguistic and ethnic descendents of Roman Dacia (though even if the continuity theory is wrong, it doesn't change the fact that the Romanians have a share in Roman cultural and linguistic heritage). The members of the Transylvanian school went to great pains to "bring the language into line" with what it should be. This marks the beginning of the period of standardization for Romanian, which itself marks an end of innocence of sorts – at least as far as language change is concerned. Before this time period, language changed naturally, as it had for centuries before. But standardization involves a conscious manipulation of language, an attempt to preserve it as is or to divert it onto a path where grammarians believe it should go – the natural partner of standardization is prescriptivism. For instance, the English rule in which one cannot end a phrase with a preposition was not drawn up because of a natural trend in English (which was running opposite this rule), but because its creator, literary critic and poet John Dryden, felt that that was how it should be, no doubt inspired by the "superiority" of Latin and French, languages which never place prepositions in a final position. Standardization is almost always based on a political motivation, be it unity, dominance, or in the case of Romanian and the Transylvanian School's members, amelioration of a language which

---

<sup>106</sup> Georgescu, p. 7

had long been written off as insignificant. The goal of the Transylvanian School was to change what some saw as a backwoods patois into a full member of the Romance family of languages.

It is important to mention that Petru Maior, the historian, fully believed that Romanian and Italian were essentially the same language. He thought that Italian had been reshaped at a later date through the works of literary masters such as Dante and Petrarch. Maior had also concluded that non-Latin elements in Romanian were of minor importance as they did not affect the structure of the language, and could and should be removed from the language, “thus purifying it of all foreign taint.”<sup>107</sup> Maior and his contemporaries felt that by removing certain undesirable (mainly lexical) elements in the language, the rightful glory of Romanian could be restored. However, one can safely say that Maior had no idea how much the structure of Romanian had actually been affected by non-Latin elements, or how, typologically, the syntax of Romanian is quite far removed from other Romance languages. Nevertheless, this connection to Italian became somewhat of an obsession for the Transylvanian School and also for later grammarians in Wallachia and Moldavia. It would affect the language mostly by opening the door for Italian and French neologisms to flood into the language, replacing many Slavonic and Greek words for abstract concepts, and introducing new words for concepts not yet represented in Romanian. Mallinson (1988) states that members of the School engaged in a sort of witch-hunt against Slavic words, replacing them with either Latin-based international neologisms from Italian or French, or with their own creations, such as *granditate* for ‘greatness,’ intended to replace *măreție*, which was based on a non-Latin word for ‘big,’ *mare*. Another example is the fabricated word for ‘war,’ *răzbel*, which

---

<sup>107</sup> Close, p. 19

attempts to folk-etymologize the Slavic loanword *război* as a Latinate word based on *bellum*. These words, and others, such as *dracone* ‘dragon,’ intended to replace the Slavic loan *zmeu*, never gained popularity and thus did not remain in common usage, to be passed on to future speakers.<sup>108</sup>

The Transylvanian School was also responsible for the first Latin-alphabet writing system designed to write the Romanian language. Romanian had, up until that time, been written using the Cyrillic alphabet. The members of the Transylvanian school recognized the need for a Latin-based alphabet if Romanian were to ever achieve the prestige and status of other Romance languages. The Transylvanian School Latin alphabet, however, was not the one used to write Romanian today. It was in fact, a very poor alphabetic system, designed to show etymological ties to Latin and Italian rather than accurately represent Romanian as it was spoken. For instance, the Romanian word for ‘man,’ *om* /om/ was represented by *homu* in an attempt to capture its origin in *homo*. Other examples are: *ventu* /vînt/ (mod. *vînt*) < *ventum* ‘wind,’ *tierra* /ʽtsarə/ (mod. *țară*) < *terram* ‘land,’ *capraa* /ʽcaprə/ (mod. *capră*) < *capram* ‘goat,’ and *monte* /ʽmunte/ (mod. *munte*) < *montem* ‘mountain.’ Their conjugation of the common verb *a avea* ‘to have’ (it also serves as an auxiliary used to form perfect tenses) simply seems ridiculous to modern sensibilities: *Am, ai, a(re), avem, aveți, au* (modern spelling) was spelled as *abiu, abi, abe, abemu, abeti, abu* to better reflect the Latin forms of *habere* -- *habeo, habes, habet, habemus, habetis, habent*.<sup>109</sup> This spelling system was completely unsuccessful, and Romanian continued to be written in the Cyrillic script until the late 19<sup>th</sup> century.

---

<sup>108</sup> Mallinson, p. 415

<sup>109</sup> Examples from Mallinson, p. 415 and Rosetti, p.139

The specific reforms of the Transylvanian school were for the most part never realized in the actual written or spoken Romanian language. This may in part be due to the fact that the rest of the Romanian peoples were politically separated from Transylvania, resulting in less direct influence. However, it was exactly the theories of the Transylvanian School that inspired a later generation of Wallachian and Moldavian grammarians to introduce their own reforms and innovations. And unlike the Transylvanian School, these writers would have a direct impact on the language itself, helping to shape Romanian into its modern form.

## **6.2 Wallachia and Moldavia**

The principalities of Wallachia and Moldavia, Romanian dominions since the Middle Ages, fell under the rule of the Ottoman Empire in the early 17<sup>th</sup> century, after a long period of Ottoman political influence and clientship. Throughout the 18<sup>th</sup> century, Wallachia and Moldavia were ruled by the Phanariotes, powerful Greeks from Istanbul who had been installed by the Ottomans. As a result, Greek influence reached its height during this time, as many of the local boyars (nobles) sought to elevate themselves by adopting Greek customs. However, the Phanariotes were generally disliked by the Romanians they ruled, and in 1821, a revolt led by the Romanian hero Tudor Vladimirescu contributed to the Ottomans' decision to remove them from power (the Ottomans had also become distrustful of the Phanariotes in the wake of a Greek nationalist movement against Turkish rule). From this point on, Wallachia and Moldavia were ruled by local Romanian boyars, and the two principalities were united into Romania in 1859. Romania did not earn its independence from the Ottoman Empire,

however, until 1878, following the Russo-Turkish War, in which Romania had fought on the side of the Russians, the winners of the conflict.

During this time period, Romanian literature flourished, and, as in Transylvania, nationalistic tendencies led to the development of a standard Romanian language. However, unlike the grammatical prescriptions of the Transylvanian school, the standardization that took place in Wallachia and Moldavia had lasting effects that actually resulted in the creation of the language which today is recognized as Modern Standard Romanian. The standardization movement that took place in Wallachia and Moldavia (the most influential authors were Wallachian) was characterized by many of the same qualities that were hallmarks of the Transylvanian School. These included an emphasis on Latin linguistic and ethnic heritage, a rejection of Slavic and other non-Latin influences, and a mass influx of loanwords from Romance languages, particularly Italian and Greek. In truth, many of the grammarians and writers working in Wallachia and Moldavia in the late 18<sup>th</sup> and early 19<sup>th</sup> centuries were influenced by the Transylvanian School, and this influence can be seen in their work. Wallachian and Moldavian grammarians would have a stronger and more lasting influence than their Transylvanian counterparts, largely due to the fact that the regions in which they were writing enjoyed virtual Romanian self-rule from 1822 onwards, and because the regions had been culturally and politically unified long before the formal unification of the two principalities in 1859. Transylvania, on the other hand, was dominated by Austrians and Hungarians who exercised a relatively tight grip on Romanian nationalism. Most importantly for Wallachia and Moldavia, from the early 19<sup>th</sup> century, Romanian was in fact the language of authority, whereas in Transylvania it was not.

Like their Transylvanian counterparts, grammarians in the principalities were convinced that Romanian and Italian were more or less the same language. Some, such as Ienăchiță Văcărescu (who predated the Transylvanian school) asserted that Romanian differed from Italian only in that it had never been standardized and enriched with loanwords directly from literary Latin.<sup>110</sup> It was Văcărescu's policy to import Italian words when no native Romanian word would suffice. By the late 18<sup>th</sup> century, French and Italian literature were very popular among the intellectual elite of the Principalities, and the influence of these works on native authors is exhibited in the large amounts of not only Italian and Latin, but also French loanwords. Romanian nationalists saw the flourishing of these cultures as sisters of their own heritage, and were eager to bring their language into its own right as a Romance idiom. Today, it is often impossible to tell whether a word originated in Latin, French, or Italian, but the large influx of Romance vocabulary has actually served to bring the language "back" into the Romance fold, albeit artificially.

Perhaps the single most important Romanian grammarian operating in Wallachia and Moldavia during this period was Ion Eliade Rădulescu, an author and publisher who wrote throughout the whole middle part of the 19<sup>th</sup> century, dying just a few years short of seeing Romania's total independence from the Ottomans. His career is characterized by two periods, one in which he emphasized the enrichment of the Romanian language using Latin and Romance languages as models and sources to shape it into a full-bodied literary language. The other, latter part of his career is characterized by an eccentric obsession with the ties that supposedly bind Italian and Romanian, and his reforms from

---

<sup>110</sup> Close, p.18-19

this period are aimed at making Romanian as much like Italian as possible, at the expense of non-Latin but long-established loanwords from Slavic, Greek, and other sources.<sup>111</sup>

Rădulescu's efforts to standardize the language with these goals in mind often led him in obscure directions. He often searched regional varieties of Romanian for forms which in his eyes more adequately resembled their Latin predecessors. And so, for instance, he proposed replacing the widespread word *știu* 'I know' with the Banat (region in the far west) form *sciu* on the grounds that this form was closer to the Latin *scio* (the *știu* form survives in all of Romania today), and shunned the Moldavian form *ghine* 'well,' in favor of the Wallachian *bine*.<sup>112</sup> The region in which the form was found did not matter – for Rădulescu it was more important that as much Latinity be preserved as possible. By 1838, Rădulescu insisted on the elimination of all "barbarisms" in the language, and went so far as to edit the works of his contemporaries, replacing non-Latin words with Italian-based neologisms.<sup>113</sup> Others, such as Barbu Mumuleanu, took a more reasonable approach, opting to keep loanwords which seemed to fit into the language and were in common usage by the ordinary people.<sup>114</sup> This became more or less the prevailing attitude, so that Romanian has retained many of its Slavic words, though sometimes they lost semantic space to Latin and Romance loanwords. This also happened to genuinely native Romanian words, so that in some cases, doublets have arisen: such as native *ceresc* 'heavenly' exists next to borrowed *celest* 'celestial,' and native *mormînt* 'tomb' has the borrowed counterpart *monument* 'monument.'<sup>115</sup>

---

<sup>111</sup> Close, p. 49

<sup>112</sup> Close, p. 65

<sup>113</sup> Close, p. 75

<sup>114</sup> Close, p. 139

<sup>115</sup> Mallinson, p. 416



Generally, it was genuine usage of a loan and not affected practice which would eventually bring an influx of Romance words into Romanian. Many of the words introduced in this way were successful, such as the words *masculin* and *feminin*, which have ousted native *bărbătesc* and *femeiesc*. Slavic, Greek, and even some native Latin words fell by the wayside as Romance words, used by popular authors, gained favor among literary circles. So, for instance, while few actually took Rădulescu's vendetta against non-Latin words seriously, his usage of Romance loans was noted and often emulated. This is one advantage that the Wallachian and Moldavian authors and grammarians had over Transylvanians – their language was the language of popular literature, government, and culture, permeating all intellectual and scientific circles in greater Romania, whereas the Transylvanians' efforts often fell upon deaf ears until their ideas later influenced those writing in the Principalities. In other words, those writing in the Principalities possessed a literary as well as political import which was lacking in the works of the Transylvanians.

Wallachian and Moldavian authors also exercised syntactic innovations which still affect the modern language. For instance, the use of the infinitive is still alive, at least in the literary language. This is largely due to its usage in the literary standard as based on the works of 19<sup>th</sup> century authors from the Principalities. Also, even more productive and widespread is the Romance formation of the passive voice. Under Slavic influence, Romanian has traditionally made use of a reflexive use of the passive, so that the phrase 'the books are stolen' is generally rendered as *se fură cărțile* (where *se* is the reflexive marker, and the subject is absent). However, gaining popularity since the 19<sup>th</sup>

century, particularly due to French influence, is the form *cărțile sînt furate*, made up of a to be auxiliary + past participle.<sup>116</sup>

On the whole, Wallachian and Moldavian innovators were successful, though maybe not to the degree that many of them had intended. Slavic words still thrive in Romanian, as do Balkan grammatical constructions, but Romance words have indeed become an integral part of the vocabulary of any Romanian speaker, regardless of educational level. Their efforts at standardizing Romanian grammar are the basis for the standard Romanian grammar of today. When all three parts of Romania were united following the First World War, the standardized Wallachian dialect, focused on the capital of Bucharest, became the norm throughout all three major regions. But the standardization practiced in the 18<sup>th</sup> and 19<sup>th</sup> centuries do mark an end of innocence – the language will no longer change on its own, without guidance, at the rapid rate that it did before. These practices mark a change brought on mostly by influence and aspiration to a higher goal – political and nationalistic at its base, yes, but for the most part a willing transformation in the speech and literature of Romanians. And it all hinges on the political importance of classification itself – to the Romanian intellectuals and later the Romanian people, the fact that Romanian is a Romance language is very important, both internally for their culture and externally for their place on the world stage.

While internal influence and prescriptivism effected a change in the Romanian language during the 18<sup>th</sup> and 19<sup>th</sup> centuries, the 20<sup>th</sup> century would mark a change of a different sort. As part of the Soviet Union, Moldova (comprised of the eastern part of historical Moldavia and Bessarabia) would become the focus of an intense linguistic engineering program, instituted from without. The Soviets introduced a policy which

---

<sup>116</sup> Mallinson, p. 416

sought to warp the dialect of Romanian spoken here, and they would use classification itself as a means to form this dialect into what Moscow wanted it to be. While Romanian grammarians from all three major regions had served to only guide the language in a certain direction (as much as many of them would have liked to change it), they never had the political clout to do exactly what they wanted. Soviet policy-makers, on the other hand, pose an interesting question: What if one were to take linguistic policy, standardization, and selective classification to the extreme, using an unlimited amount of political power to do so?

## **7 THE SOVIET UNION AND THE CREATION OF THE MOLDOVAN LANGUAGE**

The classification of the Romanian language as a Romance language has always been very important to the national consciousness of Romania. It signified ties with the languages of the rapidly-developing nations to the west which to Romanians represented a rich past and a bright future, especially when compared to the other nations of Eastern Europe, which were very much still in the clutches of feudalism well through the 19<sup>th</sup> century. Even today, Romanians are extremely proud of their linguistic heritage and its cultural implications.

However, in the 20<sup>th</sup> century, one group of Romanians discovered what it would be like to have all of this taken away from them, at least nominally. In the 20<sup>th</sup> century, Romanians in the Western part of the Ukraine became citizens of the Soviet Union, in the Moldavian Autonomous Soviet Socialist Republic (MASSR), and later, the Eastern half of the region of Moldavia was acquired by the USSR and reformed into the Moldavian Soviet Socialist Republic (MSSR), which is the modern Republic of Moldova. Soviet language policy is notorious in that the Soviets made every possible effort and used frightening amounts of force in order to construct language to their liking. The legacy of the Soviet Union in this regard is that they have provided important sociolinguistic evidence as to how far a ruling body can go to effect language change and affect language classification. Moldova was no exception to this general tendency, and in fact stands as one of the most intriguing examples, in that the Soviets tampered with a language that already had a long and well-established literary history. The policies instituted in Moldova were, however, merely one part of a much larger picture, one which spanned across two continents and almost three quarters of a century. Soviet

leaders were well aware of the importance of language to national and individual identity, and that it could be manipulated to serve as a powerful tool for the State.<sup>117</sup> They understood that classification of these languages allowed a large degree of political manipulation, and did not hesitate to use this power. Soviet policies of classification and manipulation have affected all of the languages within the former Soviet Union, collapsing some languages into others while at the same time using these very tools to drive other languages apart, so that to this day, once mutually intelligible idioms are much more foreign to one another than they once were.

### **7.1 Language Policy in the Soviet Union: From Roots-Revival to Russification**

Language policy was at the forefront of the Bolshevik Revolution in 1917. At this time, illiteracy plagued the Russian Empire – less than 30% of the population was even remotely literate, and in some regions illiteracy rates were as high as 100%.<sup>118</sup> In order to mobilize and industrialize the nation, literacy would become an unavoidable necessity. Early Soviet officials, including Vladimir Ilich Lenin, understood that without mass literacy, the changes to be brought about by the new government would mean very little to the common people themselves. Without any level of literacy, people would have a difficult if not impossible time trying to understand the lofty concepts involved in Soviet ideology. But the newly-formed Soviet Union, like the Russian Empire before it, was a multilingual nation. Many of these languages had never been written, and even fewer had been properly documented, studied, and classified. In order to develop literacy, therefore, Soviet linguists first had to decide which languages and dialects were to be developed, and how they would be brought into the modern era as full literary languages.

---

<sup>117</sup> Grenoble, p. 1: “Soviet leaders knew that language counts, that it is a crucial part of both a nation’s and an individual’s identity, and it could be manipulated to serve as a powerful tool for the State.”

<sup>118</sup> Grenoble, p. 35

This policy would have enormous implications for nationality within the Soviet Union, as one of the key components of the recognition of ethnicity in the country was the recognition of a unique language.<sup>119</sup> In many cases, Soviet officials literally had to create names for languages and ethnic groups, as the groups either did not already have names or they saw themselves as part of a larger group. Religion and cultural practice were often far more potent in one's identity than ethnicity or even language

Lenin claimed to support a policy of equality for all ethnic groups, a policy that involved "freedom and equality of languages".<sup>120</sup> He publicly advocated the right of ethnic and national self-determinism, and the idea that no language should get the status of official state language (including Russian), following the maxim "the language used to deliver the message of the Communist Party is inconsequential, compared to the message itself".<sup>121</sup> Though nationalism officially had no place in Marxism, Lenin believed that by giving these various ethnic groups a national identity they would eventually readily and freely assimilate into a higher "Soviet nationality", of which Communism would be the only significant component (in much the same way, the American nationality is not based on ethnicity, but primarily on membership in a system of common ideas). To Lenin, nationalism was a useful means to advance the proletarian cause. The Soviet tolerance of nationalism and national identity would signify that the Soviet Union was really something new, rather than just another dominating group like the Russian Empire was. He believed that by allowing these nations the self-determinism that was repressed under the tsars (it was for this very reason that many of the non-Russian nations in the Russian

---

<sup>119</sup> "official recognition of the existence of a language was, in effect, analogous to providing official recognition of a distinct ethnic group" – Grenoble, p. 20

<sup>120</sup> Grenoble, p. 35

<sup>121</sup> Grenoble, p. 41

Empire supported the Bolshevik Revolution), this would lead to enlightenment and the eventual elimination of all guise of nationality and ethnicity (including Russian), which he believed to be themselves class-constructed entities. This led to the policy of *korenizacija*, from the Russian word *koren* 'root,' meaning a return to one's roots.<sup>122</sup>

The early Soviets were faced with the daunting task of cataloguing the various languages throughout the Soviet Union, setting up grammatical and orthographic standards, and creating educational platforms from which to cure the plague of illiteracy left over from tsarist Russia, the ultimate goal being the spread of Communist doctrine. The intended result was to be a solid, Communist nation. Very few of these nationalities and ethnic groups had been documented in the first place – Soviet ethnographers and linguists had to decide which groups would be considered actual ethnic or national groups, eventually resulting in a total of 172 nationalities.<sup>123</sup> Each of these was assigned varying degrees of “importance”, which was decided by the state. Two major divisions were between those classified as *narodnost*, ‘folk / ethnic community,’ and those classified as *nacional’nost*, ‘nationality,’ the latter often having much more political and economic import. The classification of a language determined the degree to which it would be developed or in some cases, phased out (in this way it was extremely advantageous to be classified as a *nacional’nost* rather than a *narodnost*). Many smaller languages were assimilated into larger groups for economical reasons – it was simply not logistically possible to fully develop every single language within the USSR, and for educational purposes very clear lines had to be drawn between languages in order to arrive at national standards. This was in itself a difficult task since many of these

---

<sup>122</sup> Fouse, p. 89

<sup>123</sup> Grenoble, p. 39

languages were mutually intelligible with others, so that lines were often drawn based on political and economic criteria rather than linguistic. Sometimes this was done quite arbitrarily, as in the case of the Turkic languages. These languages (including Azeri, Kazakh, Uzbek, Kyrgyz, Turkmen, and Tatar among others) were spoken in more or less a dialect continuum before Soviet intervention, with very little to distinguish each adjacent dialect within a geographically contiguous area. The Soviets tried very hard to distinguish them from one another and especially from Turkish, in an attempt to avoid a potentially dangerous political solidarity. Because of this, national standards were designed to maximize differences between these languages, often by raising up obscure features in remote dialects.

From the beginning, the Soviets encountered problems concerning the scripts in which minority languages were written. Many languages had no written form at all, and many used awkward, inaccurate, Arabic-based orthographies. In the 1920s, it was decided that the Latin script was the most ideal for representing all of the non-Slavic languages within the Union,<sup>124</sup> as many officials, including Lenin, thought of Latin as a script of progress and technology (after all, the Cyrillic alphabet was seen by some as a symbol of tsarist Russia). It would also be cheaper to use in the necessary task of printing party literature, as Latin-alphabet presses were cheaper and more readily available than Cyrillic ones. Soviet officials were aware that the Arabic script, used by most of the languages in the southern USSR, would represent a significant problem, as it was directly tied to religion, namely Islam, and the non-Soviet nations in which it thrived (particularly Turkey). They decided that a Latin-based script would be instituted for these languages.

---

<sup>124</sup> Georgian and Armenian were excluded, as their writing systems and literary traditions were unique and respected as being older than Cyrillic and Slavic.



This decision was popular with all but the most religiously conservative, as the Latin script was easier to learn for the mostly Turkic and Caucasian speakers of the region, and for its utility in accurately and clearly representing the phonology of these languages much better than the Arabic script (which transcribes only consonants) did. Similar changes were made to languages which used the Old Mongolian script, so that cultural and religious (Buddhist) ties to Mongolia were at least nominally severed.

Despite this prescribed switch to the Latin alphabet, Soviet language policies were at first amazingly inclusive, especially when compared to the exclusive and Russifying policies of the tsarist Russian Empire. At this time, language policy was aimed at building national identities within each ethnic group, and the laws reflected this: by the constitution of the time, all Soviet citizens had the right to be educated in their respective native languages, and no language held official status. Soviet officials believed that by strengthening the unique ethnic identities of individual groups, literary enlightenment and self-awareness could be attained. However, it was very important that these groups remained unique – Soviet officials and language classifiers did everything in their power to avoid solidarity among ethnic blocs, and their policies reflected this by intentionally exaggerating the differences between languages that shared a common genetic origin or culture. While there was no official policy of Russification, most of the national languages in the schools and in literature received a steady stream of Russian loan words, idiom, and grammatical theory. And in the 1930s, Russification would become the norm, though it remained officially unstated.

By this time, any apparent concessions that had been given to the national languages would give way to overt Russification. It soon became apparent that Lenin's

original lofty plans for the future of the Soviet Union were not materializing. This included the idea of the unification of ethnic and linguistic minorities into a “Soviet nationality”, which was rapidly proving to be a lost cause. Many leaders, including Josef Stalin, now believed that the only way to unify the various groups was to forcefully assimilate them. Party rhetoric soon dictated this change in policy: “To make this socialist state work, in conditions of extreme centralization, there must be a maximum uniformity: one language, and if possible, one culture ... The Russian language, being the language spoken by the largest number of people in the unit, performs this role.”<sup>125</sup> Due to the failure of the World Revolution that would have been necessary for Communism to succeed in its original form (as proposed by Marx), it was clear that if the ideology of Communism would not unify these different ethnic groups, an ethnocentric policy could unify the Soviet nation. Of course, such a policy could only be applied at the expense of ethnic and linguistic minorities.

The process of Russification was mostly carried out through language and educational policies. Minority ethnic groups were allowed to retain their “national languages”, but nearly all but six of them were forced to use the Cyrillic script: Georgian and Armenian, the Finnic language Karelian (written in Latin), and the three languages of the Baltic States, Latvian, Lithuanian, and Estonian (all written in Latin). Nominally, this blanket Cyrillicization was carried out with the same intent of raising the national statuses and literacy rates of the national languages as was the Leninist use of Latin, but this time the ultimate goal would be their complete replacement by Russian. If the

---

<sup>125</sup> Pospelovsky, p. 127

Soviets were to aim for a single, national, “Soviet ethnicity”, then linguistic assimilation would ideally be the key to ethnic assimilation.<sup>126</sup>

It is important to note that the Cyrillic script chosen to represent all of these languages is based on the Russian Cyrillic script, complete with all of its peculiarities to that particular language. Letters such as **я**, **ю**, **ы**, **и**, **и**, and **ь** (/ja/, /ju/, /ɨ/, /šč/, /ts/, and a symbol which palatalizes the preceding consonant) were often brought into languages which often had no use for them because of different phonological systems. For example, **и** (/ts/) is a single phoneme in Russian, whereas in many of the languages of the Soviet Union, it was used to represent the two phoneme combination, /t/ + /s/. In cases where the languages had phonemes not found in Russian, the script was poorly adapted. For instance, the Tungus language Evenki has both a voiced bilabial /β/ and a voiced labiodental /v/ fricative as phonemes, but the Cyrillic alphabet designed for this language uses one letter, **в** (/v/ in Russian) to represent both. In some cases, one sound was covered by several symbols, such as the Evenki /ŋ/ which could be written **н**, **нг**, or **н**, depending on the word. There were many cases in which languages that were very closely related were given different orthographic systems, especially when new symbols were chosen to represent non-Russian sounds. The chart on the following page displays an assortment of sounds found in Turkic languages across the Soviet Union, arranged in geographical order from West to East. Orthographic systems do indeed differ in their spellings of the same phoneme, even within families (German uses *w* and English uses *v* to represent /v/, for example), but the differences illustrated here are not due to historical consequences, but to deliberate variegation. The Soviets used this method with moderate

---

<sup>126</sup> Dima, p. 92

**Table 2.** Chart of specific common phonemes in Turkic Cyrillic orthographies.

sound	Gagauz	Azeri	Turkmen	Uzbek	Kazakh	Chuvash	Tatar	Kyrghiz	Uyghur	Yakut	Evenki
/æ/	ă	ə	ə	--	ə	--	ə	--	ə	--	--
/ʋ,ʁ/	--	ґ	Г	Ғ	Ғ	--	Г	Г	Ғ	Б	--
/j/	й	Ј	й	й	й	й	й	й	й	й	й
/i/	и	Ы	и	и	І	и	и	и	и	и	и
/q/	--	--	К	Қ	Қ	--	К	К	Қ	--	--
/ŋ/	--	--	Ң	--	Ң	--	Ң	Ң	Ң	Н	НГ,Ң
/ø/	ö	ө	ө	ў	ө	ö	ө	ө	ө	ө	--
/u/	у	У	у	у	Ұ	у	у	у	у	у	--
/y/	ŷ	--	ү	--	ү	ŷ	ү	ү	ү	ү	--
/h/	х	Һ	Х	Ҳ	Һ	--	Һ	--	Һ	Х	--
/w/	Ы	--	Ы	--	Ы	ă	--	Ы	--	--	--
/dʒ/	Ж	Ч	Ж	Ж	--	--	Ж	Ж	Ж	--	--
/w/	В	--	--	В	У	--	В,У	--	В	--	Б

success to help break up the Turkic ethnic bloc, whose solidarity they feared greatly. The end result of this and other diversifying practices (such as the amelioration of obscure dialectal lexical and morphological forms) was that an educated Azeri speaker could not only not read publications in Turkish (written in the Latin script), but also not in Tatar or Kazakh. Although genetically very close, they were artificially made to look completely foreign.

Of course, one of the stated reasons the basic Russian orthographical system was left unchanged in these languages was so that the different ethnic groups would enjoy a certain degree of unity. But actually, the only letters shared by all of the languages were the original Russian letters, meaning that only *Russian* words could travel from language to language in their original forms. The more important advantage for Soviet officials was that it would easily facilitate the spread of new vocabulary: Russian loanwords of ideology, technology, and culture, like *sovet* “Soviet”, *bol’shevik* “Bolshevik”, *vsesojuznyj* “all-Union”, and *tovarišč* “comrade” were among many of these loanwords. Almost all loanwords from other sources first came through Russian, so that this is how words like

*kommunizm* “Communism” and *kollektivizacija* “collectivization” entered the national languages.<sup>127</sup> At first, efforts were made to calque such words or to use them with native affixes. Soviet officials touted this loaning as a sharing of ideas, a natural and positive process. However, the “sharing” only went in one direction, from Russian into the national languages. Soon all loanwords were coming directly from Russian with no modifications.

The spelling of loanwords soon became almost as important an issue as the borrowings themselves. The “Common Rule” decree of the 1940s mandated that all Russian loanwords as well as international (mostly Latinate) borrowings must be written and spelled as in Russian. The rule conveniently overlooked the fact that the Russian spelling reflected the Russian pronunciation, often impossible in the phonological systems of these languages. This resulted in gross discrepancies between spelling and pronunciation: In Yakut, “table” is spelled as *стол* to reflect the Russian pronunciation /stol/, though it is pronounced /osto:l/. In Bashkir, “number, date” is spelled *число* (/čislo/ in Russian), but pronounced /sisal/.<sup>128</sup> The mandated spelling conventions also ignored morphological elements of the national languages, so that if it appeared at all, morphological orthography was irregular and confusing. In practice, it seemed as if these languages had to employ two different spelling systems, a Russian one and a native one. This hindered the ability of children to learn to read and write in their native language, even as they acquired their second language, Russian. Though the Common Rule was officially abolished in 1950, damage had in most cases already been done, and the orthographic rules were only partially revoked.

---

<sup>127</sup> Grenoble, pp. 52-53

<sup>128</sup> Grenoble, p. 53

In addition to deliberately tampering with phonology and morphology, Soviet educators attempted to force Russian syntactical and idiomatic structures into the national languages. Literal translations from Russian sources were often claimed to be the correct structure of the recipient language, and grammar taught in schools was often Russian grammar and sentence structure with the words and other morphemes replaced by native ones (this is in a way similar to prescriptive grammar in English, which goes out of its way to emulate Latin syntactic conventions). Idiom was also taken directly from Russian, regardless of the fact that Russian idiomatic expressions often came from a different cultural context than that of the recipient language. This problem was especially potent in the East Slavic languages closely related to Russian, Belarussian and Ukrainian, which had been stigmatized for centuries as lesser, non-standard dialects of Russian. This meant that education in and development of these languages was often slanted towards “reuniting” them with standard Russian.

As a result of the confusion and difficulty involved in education in the national languages, it was all too easy for students to simply opt to learn and use Russian instead, since the native languages, although officially respected, were becoming increasingly unnecessary and stigmatized. There were many obvious professional and social advantages to learning Russian, whereas the purpose of developing one’s native or heritage language was unclear at best. And while education in the national languages was required by the policies set forth by Lenin, educational materials were often poor and hard to come by, as opposed to Russian educational materials which abounded from Moscow. Russian had also from earliest times in the USSR been seen as a *lingua franca*, a means of intercommunication between the various peoples of the Soviet Union. All in

all, there were serious disadvantages to not knowing Russian, but no real disadvantages to *only* knowing Russian. This had an enormous impact on students, who realized that the only viable avenue to success was through proficiency in Russian and not in their own native languages. Although bilingualism was officially valued in the Soviet Union, it was the one-way bilingualism between the national languages and Russian that really mattered.

In the years following the Second World War, Russification policies became increasingly stringent. Nikita Khrushchev openly questioned the need for any education in the national languages. Russification was often acknowledged, but characterized as a conscious choice: “the will of the people to move forward in progress”<sup>129</sup>. Steps closer and closer to complete Russification were made, officially excused, and never questioned. Education reforms in the late 1950s made education mandatory for Russian and optional for the national languages, prompting many schools to cut their already weak native language programs. And by the late 1970s, many of the national language programs were completely abandoned except as cultural enrichment electives, and almost all education in other subjects was done in Russian.<sup>130</sup> The 1977 Constitution made the policy of Russification quite clear: where the 1936 Constitution gave Soviet citizens “the *right* of school instruction in one’s native language”, the 1977 Constitution assures the *opportunity*.<sup>131</sup> The 1936 Constitution demanded such instruction, whereas the 1977 Constitution merely entertained the possibility of access. At this point, at least within

---

<sup>129</sup> Bodrogligeti

<sup>130</sup> In the 1930s, mathematics and sciences were taught in at least 65 different languages, but in the 1980s they were taught in only 32 – Grenoble, p. 60.

<sup>131</sup> Grenoble, p. 58

party rhetoric, the idea of a unified Soviet nationality was becoming a reality, though not in quite the same way Lenin had envisioned it.

By the late 1980s, as Soviet power over its constituent nations began to wane, many ethnic and linguistic groups began taking chances at declaring their cultural independence from the Soviet Russians after a long relative silence. It became apparent that the idea of an “ultimate fusion into some nonethnic and nonnational community” did not sit well with most of the peoples within the USSR, and that any apparent willingness to become such a community was only the result of a forced educational system. Soon, all of the SSRs within the USSR began to take liberties with their own national languages, and, beginning with Moldova and the Baltic States, they one by one declared their national languages as state languages. In 1990 the government in Moscow responded with a statement declaring Russian the *official language* of the USSR for the first time in its history. Of course, they could not stem the already sweeping tide of social change, and when the Soviet Union fell apart in 1991, the national languages were well on their way to reclaiming what they had lost.

## **7.2 The Construction of the Moldovan Language**

The modern Republic of Moldova is the descendant of the Moldavian Soviet Socialist Republic, a member state of the Soviet Union. The historical-geographical name of this region between the Prut and Dniester Rivers is Bessarabia. During much of the time when Romania was composed of three different political entities, Bessarabia was simply the eastern half of the principality of Moldavia, a vassal state within the Ottoman Empire. However, in 1812, after the end of the Russo-Turkish War, this region was ceded to the Russian Empire. For a brief time between the two World Wars, the region was



again united with Romania (taking advantage of the unstable political situation in the Russian Empire in the wake of the Bolshevik Revolution). But in 1940, Bessarabia was annexed by the Soviet Union, who created the Moldavian SSR in its place.

“Moldovan” is a Romance language with about 2.6 million speakers in Moldova and in the Western Ukraine. It is the only Romance language that was officially spoken in the USSR. The Moldovan language is undeniably linguistically identical to the Romanian language. It could hardly even be considered a regional dialect, any more than “Texan” could be considered a regional dialect of American English.<sup>132</sup> If anything, it is an accent of the regional dialect of Romanian spoken in Northeastern Romania (even this is little more than an accent). But this means very little, as regional dialects in Romanian depart from the Bucharest-based standard in only minor ways, such as differences in pronunciation and a few differences in vocabulary (much more like regional dialects in the United States than in other European countries, like Germany and Italy). Most of the differences that *do* exist today are the direct result of Soviet language policy in the Moldavian SSR throughout the 20<sup>th</sup> century, and they usually consist of the use of a Russian or Ukrainian loanword in the place of a Romanian word. But though the languages remain identical to linguists, the classification of Moldovan as Romanian is still a very hot issue in the politics of the region. The question as to whether or not they *are* the same language must remain officially unanswered, because to admit that they are the same language gives credence to the idea that the two nations should somehow be reunited. In Romania, this view is favored by many, as it is in Moldova. However, there is also a large faction in Romania which wants to have nothing to do with Moldova, the poorest and most unstable country in Europe. Similarly, in Moldova, a large part of the

---

<sup>132</sup> Dima, p. 95

population would like to remain distinct from Romania, as they feel that their different political circumstances for the past century have divided the two peoples enough so that they are now two distinct cultures. The debate will not be resolved any time soon, and as it stands now, the official language of Moldova is Moldovan and not Romanian, an issue of language designation which makes a very strong political statement that transcends any classification a linguist would make.

The character of the Moldovan language is the product of Soviet language policy that governed the language for roughly 65 years. While Bessarabia had been a part of the Russian Empire since 1812, the inhabitants of the region were recognized as Romanians who spoke Romanian. No attempt was made to develop this language within the Russian Empire, and the Romanian of Bessarabia remained in linguistic continuity with the Romanian of Western Moldavia despite the political and geographical boundary of the Prut River. The people of this region were simply Romanian speakers living within the borders of the Empire, just as there were Polish, Czech, and even German speakers doing the same. Literacy was not an issue, as most of these people were illiterate, as was the majority of the Russian Empire. When the Soviet Union arose, however, things changed. As part of the general language policy to ameliorate the national languages of the Soviet Union, policy-makers brought literacy and linguistic development to this region as well. However, what they attempted to develop would not be the Romanian language, but a language which they insisted had a separate identity: Moldavian. “Moldavia” and “Moldavian” reflects the Soviet terms *Moldavija* and *moldavskij* – this term was used throughout the Soviet period, entailing the Soviet viewpoint on the language, whereas today the terms “Moldova” and “Moldovan” are used, reflecting the Romanian *Moldova*

and *moldovenească* (confusingly, the common name for the principality which existed before the modern period, as well as for the modern region in Eastern Romanian, is also “Moldavia” in English).

When the Soviets first applied linguistic policies to the Moldavian language, the Moldavian SSR did not exist, as Bessarabia was at this time a part of Romania. However, in 1924, the Moldavian Autonomous Soviet Socialist Republic was created within the boundaries of the Ukraine, in an area which bordered Bessarabia and in which Romanian was spoken by a large part of the population.



**Figure 6.** Romania and the Moldavian ASSR (map from *Wikipedia*).

Even though the MASSR only had about a 32% Romanian population, it was here that the Soviets began to construct a Moldavian ethnic identity and engineer a Moldavian language. The intention of the Soviets was to create a sense of nationalism among the “Moldavian” population, specifically one which was different from the Romanians to the West. Moldavian national identity, separate from Romanian, would be one of the reasons

given as justification for the annexation of Bessarabia by the USSR in 1940. If they could prove that the Moldavians were a separate ethnic group, then such an annexation could and would be touted as a liberation rather than a hostile takeover. In order to do this, Soviet linguists worked hard to find and to create “scientific” reasons to classify Moldavian as a separate language from Romanian and a separate culture and people from the Romanians.

In the time period following the creation of the MASSR, Soviet officials released vast amounts of “nationalist” propaganda to raise ethnic awareness and stress the differences between the Romanian and Moldavian languages and cultures. Schools, now ubiquitous after the Soviet Union had launched its literacy campaign, taught the Moldavian language and stressed a view of Moldavian history in which the people participated in a cultural history apart from that of the Romanians. All of these claims were entirely fabricated or exaggerated. Romanian was decried as the language of the “bourgeois feudal ruling class”, while, according to Soviet linguists, Moldavian was a “language of the people”.<sup>133</sup> Ties between Moldavian and Slavic languages and culture were stressed, but much of this came from the fact that the Moldavian ASSR was primarily a Slavic region with *some* Romanians living in it, rather than the official view which held that it was a region of mixed Slavic and Moldavian character, essentially one people speaking two different languages, Moldavian and Ukrainian. In this view, the Moldavians were, in essence, Slavs who had acquired a Romance dialect.

Despite the artificial attempts at Moldavian nationalism, the language continued to be written in the Latin script following the Romanian literary standard until 1938. In this year, tremendous changes were made to the linguistic situation in the Moldavian

---

<sup>133</sup> Fouse, p. 89

ASSR. The most visible area of change was in the orthography of the language. Though many European minorities living in the Western part of the USSR (which is what the Moldavians were, in essence), such as Poles, Germans, Yiddish-speaking Jews, and others, were allowed to retain their literary standards, the situation in Moldavia was different for one major reason, and that reason was Bessarabia. Bessarabia had formerly been Russian territory, and the Soviet Union sought to reclaim it. They believed that Romania was occupying lands which were rightfully theirs, and by “proving” that the Moldavian ethnicity and language of the MASSR and of Bessarabia was separate from Romanian, they could use this to justify the “liberation” of these peoples from their Romanian occupiers. This is one of the major reasons the literary standard of Moldavian in all literature and in schools was differentiated from Romanian to an extreme degree wherever possible. The idea was to create a Moldavian identity while at the same time preventing the formation of a larger Romanian identity.

Orthography became a key component in this change, as it eliminated all orthographic ties between Romanian and the Moldavian language, making Romanian and Moldavian look, as dissimilar as possible, while creating a very significant superficial similarity between Moldavian and Slavic languages like Ukrainian and Russian. A new alphabet and orthographical rules were developed for the Moldavian language. Instead of simply using the old Romanian Cyrillic alphabet (used until the latter half of the 19<sup>th</sup> century), or even the Bulgarian alphabet (which is closer historically, structurally, and phonologically to Romanian), the Soviets patterned the Moldavian alphabet on the Russian version of the Cyrillic script, as they did with the other national languages. Often, this alphabet failed to represent certain morphological alterations which are clear

in the Romanian Latin alphabet: For instance, in Romanian, the relationship between the masculine, feminine, and plural versions of ‘dry’ is apparent in the orthography: *sec*, *seacă*, *seci*. In the Cyrillic orthography, the relationship is a bit less obvious: *сек*, *сякэ*, *сеч* (It should be noted that the phonologically distinct diphthongs *ia* (/ia/) and *ea* (/ea/) are represented by the same grapheme in the Cyrillic version). In Moldavia, the end result of the change to Cyrillic was that a child learning to read would learn Moldovan, but would have no access to Romanian without learning an entirely new alphabet.

Any attempt to resist this change within the Moldavian ASSR was met with severe punishment, often in the form of Stalinist purging – namely, death – as an enemy of the State or agent of Romanian imperialism. The Soviets claimed that any use of the Latin alphabet was an attempt at “Romanianizing” Moldavian (and thus, Romanianizing something which was already Romanian).<sup>134</sup>

In 1940, Bessarabia was finally annexed by the Soviet Union from Romania, creating the Moldavian Soviet Socialist Republic, a full-fledged member state of the USSR. The Old Moldavian ASSR was divided in half, the western side becoming part of the Moldavian SSR, and the Eastern side being assimilated into the Ukrainian SSR. (see map). As opposed to the Moldavian ASSR, which had a mostly Slavic population, the new Moldavian SSR had a majority Moldavian (Romanian) population because of the inclusion of Bessarabia. The rest of history of language policy in the Moldavian SSR until 1989 is replete with Soviet engineering of the language.

The basic principle touted by the Soviets concerning the Moldavians was that they were originally a Slavic people “colonized” by Romanians (this is what was taught in schools as well – Romanians were long criticized in Moldavian textbooks as aggressors

---

<sup>134</sup> Bruchis, p. 49



**Figure 7.** Modern Moldova (Map from the *CIA World Factbook*, 2006).

and colonizers).<sup>135</sup> Soviet language policy claimed that the Moldavian language is an East Romance language similar to Romanian, but with a heavy Slavic substrate and adstrate. The policy always held that Romanians and Moldavians were ethnically distinct, and this worked in favor of twisting the Moldavian language in order to bring it more in line with “fraternal” peoples, namely, Slavs.

The most extreme version of the theory, one which was advocated during the Stalinist era, has the Moldavian language as a Slavic language which was overlaid with a heavy superstrate of Romanian. It was during this era that those calling themselves

<sup>135</sup> Dima, p. 122

“purists” began the process of “purging the Moldavian language of Gallicized Romanian words introduced by enemies of the Moldavian people”.<sup>136</sup> These words included many of the words introduced in the 19<sup>th</sup> century by Romanian writers and grammarians, a large group of neologisms based on Latinate roots and loanwords from Italian and French. The purists also made use of as many loanwords as possible from Russian and Ukrainian, as well as creating “Moldavian” neologisms not only from native parts already available in the language (as opposed to Romanian neologisms which were made directly from Latin, not Romanian), but also from Slavic roots.

The purists also managed to effectively ban all Romanian literature from the Moldavian SSR and promote “native Moldavian” literature, which had to be approved by Moscow, so that, in effect, the only source of literature in Moldavia came from Russia itself in one way or another. Any writer caught writing in the Romanian literary language (opting not to use the artificial Soviet innovations) was publicly reprimanded as “trying to drag their aristocratic drawing-room words into the language.”<sup>137</sup> All of this had immense implications for the development of Moldavian as a language. For one thing, Soviet officials now controlled education as they did in other regions of the Soviet Union. The result was the creation of a new literary standard for the language, a standard free of “Romanianizations.” This basically meant that they imposed Slavic grammar and idiom into the classrooms and into the editing of current literature. The Soviets had full control of any literature before it was released into the SSR, and they used this to their advantage. Many words of Latinate origin, particularly those referring to abstract concepts or aspects of political life, were replaced wholesale either by Russian words subject to the Common

---

<sup>136</sup> Bruchis, p. 61

<sup>137</sup> Dima, p. 97



Rule (many writers were not even native Moldovans, and freely and deliberately placed Russian and Ukrainian words into their publications: *soiuz* for *uniune* ‘union’ and *shrift* for *scris* ‘script,’ *načal’nik* for *director/președinte* ‘head, director,’ *zadachă* for *treabă* ‘task’)<sup>138</sup> or by fabricated words made up from Slavic roots. Another common tactic was to revive old Romanian Slavic borrowings that had fallen into disuse. So, for instance, the Slavic loanwords *norod* and *ukaz*, ‘people’ and ‘decree’, replaced *popor* and *decret*.<sup>139</sup> In addition, all foreign place names were to be written and pronounced as in Russian (Romanian *Olanda*, *Polonia*, *Roma* ‘The Netherlands,’ ‘Poland,’ ‘Rome’ were replaced by forms based on the Russian *Golandija*, *Pol’ša*, *Rim*)<sup>140</sup> There were even cases where if a native word sounded too similar to a taboo word in Russian, it was removed or replaced. The words *hui* and *huidui* (in Cyrillic *xyu* and *xyudyu*), ‘to buzz’ and ‘to hiss’ were banned from print because they resemble the Russian word *xyŭ* (/xuj/), ‘cock or dick,’ one of the most vulgar words in the Russian language. Similarly, the second person singular form of the verb *a putea* ‘to be able,’ *poți*, was discouraged because it resembled the Yiddish word *putz*.<sup>141</sup> Overall, what was described by officials as a “campaign against cluttering the language with words incomprehensible to the people [that is, mostly literary and abstract words created in the past century or imported from western sources]”<sup>142</sup> did just that – it infused unfamiliar words and concepts into a language which already had familiar words for these things. Nationals as well as the occasional visitor from Romania would often complain that the language as seen in the newspapers, especially during the years of the Moldavian ASSR, was fraught with a cacophonous

---

<sup>138</sup> Bruchis, p. 75

<sup>139</sup> Dima, p. 98

<sup>140</sup> Bruchis, p. 76

<sup>141</sup> Bruchis, p. 101

<sup>142</sup> Bruchis, p. 85

mixture of Russian and Ukrainian words and sentence structure with the native Moldavian. Of course, such complaints meant little to and were often ignored by the government itself. After all, they had an agenda of their own.

Blatant attempts at manipulating the language to facilitate Russification were met with much opposition by scholars and writers in the Moldavian ASSR and later in the Moldavian SSR, who would often use over-Russified forms in their work in order to ridicule the state of the language.<sup>143</sup> As Soviet power weakened and policies became less stringent, cultural ties with Romania once again became an important nationalistic issue in Moldavia. Writers openly declared that Moldavian was a dialect of Romanian (if not simply Romanian) and stressed the importance of reuniting the two languages. The 1980s saw a boom in Moldavian literature, as well as a rise in the education of Moldavian in schools (which was neglected under Krushchev and Brezhnev). In 1988, Republic officials suggested that Moldavian be named the official language of the SSR, setting a precedent within the Soviet Union, eventually resulting in the declaration of Russian as *the* state language of the USSR. In 1989, the official script was officially changed to Latin, though with a Moldavian orthography that was different from Romanian. In that same year, the law banning Romanian literature was repealed, and Romanian books, magazines, and other publications flooded the country. When the Soviet Union officially dissolved, “Moldavian” changed to “Moldovan” and the Romanian orthography was instituted in full. Despite an international linguistic study and the conclusion that Moldovan is, in fact, Romanian, Moldovan is still considered a distinct language in the nation’s constitution, and the debate over the language’s identity continues.<sup>144</sup> In 1994,

---

<sup>143</sup> Bruchis, p. 76

<sup>144</sup> Fouse, p. 95

there was a movement to change the name of the language in the constitution to “Romanian”, but this was soundly defeated.<sup>145</sup>

Today, there is a large element within Moldova which believes that Russian should remain a co-official language, due to two major reasons: there is a large population of non-ethnic Moldovans within the country who do not speak the language but are proficient in Russian, as are many adult Moldovans. Perhaps one quarter of the population is Russian or Ukrainian, and this is even higher in urban areas and in the extreme Eastern Trans-Dniester region, which is home to an active Slavic separatist movement (this region is mostly composed of the Western half of the Moldavian ASSR, which had been carved out of the Ukraine)<sup>146</sup> Also, Moldova still remains more or less within the Russian economic sphere of influence as a member of the Commonwealth of Independent States. Many of the Russians arrived in the country as a result of the Soviet policy of ethnic dilution over the past decades which gave concessions to ethnic Russians who moved outside of Russia proper. While there are many practical advantages to giving Russian such a status, the symbolism does not rest well with most who consider themselves Moldovan. Any plans for compulsory education in Russian or recognition of Russian as co-official with Moldovan have been strongly opposed and protested – as recently as 2002 there were mass protests in the streets of the capital city Chişinău against the Communist government (Communists were voted back into the government in 2001, and still retain a majority as of 2005 – they consistently support strong ties to the former Soviet Union, and shy away from relations with Romania and the West), which had decided to institute plans to encourage the learning of Russian and to once again

---

<sup>145</sup> Jeffries, p. 28

<sup>146</sup> Jeffries, p. 326

rewrite textbooks to deemphasize the relationship between Romanian and Moldova. The government was forced to relinquish their plans. A more important linguistic issue today is whether or not Moldovan should be considered the same language as Romanian, or if the gap between the two is now too deep for the two dialects to be completely unified. Soviet language policies have indeed gotten into the structure of Moldovan, leaving it full of Russian and Ukrainian lexical items, as well as certain idiomatic expressions and grammatical constructions. But in the past decade or so, the artificial effects of the Soviet policies have been wearing off as Moldovan resembles Romanian more and more. This question is a major component of a much larger political issue – that is, total unification with Romania. Many Moldovans would like to unite with their ethnic and linguistic neighbors, but a significant element, especially within the Communist government, would like to see Moldova and Moldovan kept separate from the democratically-minded Romania.

## 8 CONCLUSION

Language designation is, and has always been, inextricably tied to politics and matters of national identity. For no other reason than this do we designate Danish, Swedish, and Norwegian as separate languages, yet Low German, High German, Swabian, Bavarian, and Swiss German as dialects of the same language. A brief survey into the former group using the tools of linguistic classification would prove that they have much more in common with each other than do the ‘dialects’ in the latter group (and even more perplexing is the fact that linguistically, Low German is much closer to Dutch than it is to standard German). Certainly, the classification of Romanian as a Romance language signifies much more than a simple linguistic affiliation – it represents an entire Romance history and culture that Romanians can claim as their own. Similarly, the separation of Moldovan from Romanian is meant to deny this inheritance. While the two “languages” are mutually intelligible and share the exact same affiliations through all three types of classification: genetic, areal, and typological, they remain designated as two separate languages.

The very nature of language designation is necessarily political, as is any demographic or ethnic designation system. When one decides that one group is separate from another group, one makes a political statement, in effect drawing a dividing line between “us” and “them.” Sometimes, the line is not drawn, even when intelligibility wanes between two speech varieties designated as one language. Such is the case in China, where at least eight full-fledged mutually unintelligible languages are designated as ‘Chinese’ despite their differences which can be demonstrated by both genealogical and typological classification methods (this is facilitated in part by the use of a common

logographic writing system for all of the ‘dialects’). And sometimes, lines may be deliberately erased or weakened, as the Romanian intellectuals and prescriptivists did when they sought to bring Romanian into line with other Romance languages through the use of loanwords, neologisms, and orthographic reforms. Finally, lines may be drawn and fortified where before there were no lines at all – Soviet officials tried to create (and succeeded, to a point) a new “language” out of what was before little more than a regional accent. The truth is that, linguistically speaking, these lines between languages and between dialects do not always necessarily exist in language itself.

In the case of Moldovan, there is very little reason for a linguist to classify it as a separate language from Romanian. In all ways, the two should be and are classified as the same language, yet on paper they remain distinct, especially in Moldova where the constitution names Moldovan as the official language. The example of Moldovan indicates that there is, in fact, a difference between linguistic classification, a scientific process based on empirical evidence; and the familiar brand of language designation that occurs in the real world, where nothing is neutral, but everything is colored by political and nationalistic claims. Within this framework, Moldovan will continue to be designated as a distinct language, no matter what can be proven by linguists using the tools of linguistic classification. In the same way, Serbian and Croatian will be designated as distinct while Cantonese and Mandarin remain a single language in the eyes of much of the world. Such designations nominally use many of the practices employed by linguists, but they are actually not much more than products of a political and nationalistic consciousness.

The question remains as to whether or not linguistic classification can be reconciled with the politics of nationalism and ethnicity. The answer, I believe, is no, they cannot be entirely reconciled. Linguists, as scientists are compelled and expected to view the world neutrally, with recourse to empirical evidence and scientific fact, ignoring political and social bias at all costs. For a linguist to make a classification or a theory of language history based on a nationalistic agenda would be viewed as a departure from, or even a betrayal of, the scientific principles which guide the field. This is why the claims of the supporters of the Theory of Daco-Romanian Continuity cannot be sustained in light of competing linguistic evidence: they are merely thinly-veiled attempts at using linguistics and archaeology to serve political purposes.

On the other hand, languages, like the people who use them, exist within the framework of the world itself, which is complete with competing political, national, religious, and personal points of view. While linguists may have no difficulty making a classification using only empirical classification methods, the world at large will never be able to organize language with such sterile indifference. If anything, it is only natural that people and governments typically classify language using not only (and maybe not at all) linguistic classification methods, but also nationalism, political sentiments, and individual experience. The truth is that to many people, it is in fact nationalism and ethnic loyalty that are most important, not science. Because of this, proper linguistic classification can never be fully reconciled with national politics, though some of its methods and tools will inevitably be used in situations where the motives for classification are primarily non-linguistic, as they were by the Soviet Union and by Romanian prescriptivists. There is no authority which can prescribe a purely linguistic approach to classification, and so,

linguistic classification must therefore be recognized as separate from common language designation. While linguistic classification does play a role in politics and nationalism, it must itself remain independent of these real-world forces in order to maintain its own principles.



## REFERENCES

- Berdichevsky, Norman. (2004) *Nations, Language, and Citizenship*. Jefferson, NC: McFarland and Company.
- Bobango, Gerald J. (1979) *The Emergence of the Romanian National State*. New York: Columbia University Press.
- Bodea, Cornelia and Virgil Căndeă. (1982) *Transylvania in the History of the Romanians*. Boulder, CO: East European Monographs / New York: Columbia University Press.
- Bruchis, Michael. (1982) *One Step Back, Two Steps Forward: On the Language Policy of the Communist Party of the Soviet Union in the National Republics (Moldavian: A Look Back, a Survey, and Perspectives, 1924-1980)*. Boulder, CO: East European Monographs / New York: Columbia University Press.
- Bruchis, Michael. (1984) *Nations – Nationalities – People: A Study of the Nationalities Policy of the Communist Party in Soviet Moldavia*. Boulder, CO: East European Monographs / New York: Columbia University Press.
- Bruchis, Michael. (1988) *The USSR: Language and Realities: Nations, Leaders, and Scholars*. Boulder, CO: East European Monographs / New York: Columbia University Press.
- Campbell, Lyle. (1997) “Typological and Areal Issues in Reconstruction”, in: Jacek Fisiak (editor), *Linguistic Reconstruction and Typology*. Berlin / New York: Mouton de Gruyter, pp. 49-72.
- Campbell, Lyle. (1998) *Historical Linguistics: An Introduction*. Cambridge, MA: The MIT Press.
- Campbell, Lyle. (in press) “Areal Linguistics”, in: Keith Brown (editor). *Encyclopedia of Language and Linguistics (2<sup>nd</sup> Edition)*. Oxford: Elsevier.
- Castellan, Georges. (1989) *A History of the Romanians*, translated by Nicholas Bradley. Boulder, CO: East European Monographs / New York: Columbia University Press.
- Close, Elizabeth. (1974) *The Development of Modern Rumanian: Linguistic Theory and Practice in Muntenia, 1821-1838*. London: Oxford University Press.
- Comrie, Bernard. (1981, 1989) *Language Universals and Linguistic Typology*. Chicago: The University of Chicago Press.

- Comrie, Bernard. (1993) "Typology and reconstruction", in: Charles Jones (editor), *Historical Linguistics: Problems and Perspectives*. London / New York: Longman, pp. 74-97
- Cristo-Loveanu, Elie. (1962) *The Romanian Language*. New York: Columbia University, self-published.
- Croft, William, Keith Denning, and Suzanne Kemmer. (1990) "Typology and Diachrony in the Work of Joseph H. Greenberg", in: William Croft, Keith Denning, and Suzanne Kemmer (editors), *Studies in Typology and Diachrony: Papers Presented to Joseph H. Greenberg on His 75<sup>th</sup> Birthday*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp. ix-xviii.
- De Bray, R.G.A. (1980) *Guide to the South Slavonic Languages (Guide to the Slavonic Languages, Third Edition, Revised And Expanded, Part 1)*. Columbus, OH: Slavica Publishers.
- Deletant, Dennis and Yvonne Alexandrescu. (1992, 2003) *Teach Yourself Romanian*. Chicago: Contemporary Books.
- Dima, Nicholas. (1991) *From Moldavia to Moldova*. Boulder, CO: East European Monographs / New York: Columbia University Press.
- Dimitrescu, Florica. (1967) *Introducere în fonetica istorică a limbii române*. Bucharest: Editura Științifică.
- Du Nay, André. (1996) *The Origins of the Rumanians: The Early History of the Rumanian Language*. Toronto / Buffalo: Matthias Corvinus Publishing.
- Dzyuba, Ivan. (1968) *Internationalism or Russification?: A Study in the Soviet Nationalities Problem*. London: The Camelot Press.
- Fennel, Barbara A. (2001) *A History of English: A Sociolinguistic Approach*. Oxford: Blackwell Publishers.
- Friedman, Victor A. (1999) *Linguistic Emblems and Emblematic Languages: On Language as a Flag in the Balkans*. Columbus, OH: Ohio State University Department of Slavic and East European Languages and Literatures.
- Forston, Benjamin W. (2004) *Indo-European Language and Culture: An Introduction*. Malden, MA / Oxford: Blackwell Publishing.
- Fouse, Gary C. (2000) *The Languages of the Former Soviet Republics: Their History and Development*. Lanham, MD: University Press of America.

- Gamkrelidze, Thomas V. (1997), "Language Typology and Linguistic Reconstruction", in: Jacek Fisiak (editor), *Linguistic Reconstruction and Typology*. Berlin / New York: Mouton de Gruyter, pp. 25-48.
- General Directorate of the State Archives in the Socialist Republic of Romania. (1979, 1980) *Foreign Sources on the Romanians*, translated by Delia Răzdolescu. Bucharest: Direcția generală a Arhivelor Statului.
- Georgescu, Vlad. (1984, English edition 1991) *The Romanians: A History*. translated by Alexandra Bley-Vroman. Columbus, OH: Ohio State University Press.
- Gilărevskij, R.S., and V. S. Grivnin. (1964) *Opredelitel' jazykov mira po pis'mennostām*. Moscow: Izdatel'stvo "Nauka."
- Goyaverts, D.L. (1975) *Present-Day Historical and Comparative Linguistics: An Introductory Guide to Theory and Method*. Ghent / Antwerp: E. Story – Scientia P.V.B.A. Scientific Publishers.
- Greenberg, Joseph H. (1963) "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements", in: Keith Denning and Suzanna Kemmer (editors, 1990), *On Language: Selected Writings of Joseph H. Greenberg*. Stanford, CA: Stanford University Press, pp. 40-70.
- Grenoble, Lenore A. (2003) *Language Policy in the Soviet Union*. Dordrecht / Boston / London: Kluwer Academic Publishers.
- Gvozdanović, Jadranka. (1997) "Typology and Evaluation of Linguistic Reconstruction", in: Jacek Fisiak (editor), *Linguistic Reconstruction and Typology*. Berlin / New York: Mouton de Gruyter, pp. 103-122.
- Harris, Martin. (1988) "The Romance Languages", in: Martin Harris and Nigel Vincent (editors), *The Romance Languages*. New York: Oxford University Press, pp. 1-25.
- Hitchins, Keith. (1996) *The Romanians, 1774-1866*. Oxford: Clarendon Press.
- Hupchick, Dennis P. and Harold E. Cox. (2001) *The Palgrave Concise Historical Atlas of the Balkans*. New York: Palgrave.
- Illyés, Elemér. (1988) *Ethnic Continuity in the Carpatho-Danubian Area*. Boulder, CO: East European Monographs / New York: Columbia University Press.
- Jelavich, Barbara. (1984) *Russia and the Formation of the Romanian National State, 1821-1878*. Cambridge: Cambridge University Press.

- Joseph, Brian D. (1983) *The Synchrony and Diachrony of the Balkan Infinitive: A Study in Areal, General, and Historical Linguistics*. Cambridge: Cambridge University Press.
- Koerner, E.F. Konrad. (1997) "Toward a History of Linguistic Typology", in: Jacek Fisiak (editor), *Linguistic Reconstruction and Typology*. Berlin / New York: Mouton de Gruyter, pp. 1-24.
- Krupenikov, I.A., D.A. Mirskij, and M.M. Padul. (1970) *Sovetskij Sojuz: Moldavija*. Moskow: Izdatel'stvo Mysl'.
- Lass, Roger. (1993) "How real(ist) are reconstructions?", in: Charles Jones (editor), *Historical Linguistics: Problems and Perspectives*. London / New York: Longman, pp. 156-189.
- Lehiste, Ilse. (1988) *Lectures on Language Contact*. Cambridge, MA / London: The MIT Press.
- Lenin, Vladimir Il'ich. (1983, editor unnamed) *Lenin on Language*. Moscow: Raduga Publishers.
- Mallinson, Graham and Barry J. Blake. (1981) *Language Typology: Cross-Linguistic Studies in Syntax*. Amsterdam / New York/ Oxford: North-Holland Publishing Company.
- Mallinson, Graham. (1988) "Rumanian", in: Martin Harris and Nigel Vincent (editors), *The Romance Languages*. New York: Oxford University Press, pp. 391-419.
- Mann, Stuart E. (1984) *An Indo-European Comparative Dictionary, Volume I: A-M, Fascicle 1-2*. Hamburg: Helmut Buske Verlag.
- McWhorter, John. (2001) *The Power of Babel: A Natural History of Language*. New York: Times Books, Henry Holt and Company.
- Mitrasca, Marcel. (2002) *Moldova: A Romanian Province Under Russian Rule: Diplomatic History from the Archives of the Great Powers*. New York: Algora Publishing.
- Mitu, Sorin. (1997, 2001) *National Identity of Romanians in Transylvania*. translated by Sorana Corneanu. Budapest: Central European University Press.
- Musaev, K. M. (1965) *Alfavity jazykov narodov SSSR*. Moscow: Izdatel'stvo "Nauka".
- Ornea, Z. (1999) *The Romanian Extreme Right: The Nineteen Thirties*. translated by Eugenia Maria Popescu. Boulder, CO: East European Monographs / New York: Columbia University Press.

- Oțetea, Andrei. (1985) *A Concise History of Romania*. English edition edited by Andrew McKenzie. London: Robert Hale.
- Pei, Mario. (1976) *The Story of Latin and the Romance Languages*. New York: Harper and Row.
- Pop, Ioan-Aurel. (1996) *Romanians and Hungarians from the 9<sup>th</sup> to the 14<sup>th</sup> Century: The Genesis of the Transylvanian Medieval State*. translated by Cristina Felea and Veronica Csorvósi. Cluj-Napoca: Fundația Culturală Română.
- Pop, Ioan-Aurel. (1999) *Romanians and Romania*. Translated by Alexandra Dimitrescu and Ioana Nan. Boulder, CO: East European Monographs / New York: Columbia University Press.
- Pop, Sever. (1948) *Grammaire Roumaine*. Bern: Éditions A. Francke S.A.
- Posner, Rebecca. (1996) *The Romance Languages*. Cambridge: Cambridge University Press.
- Rivero, María Luisa and Angela Ralli (editors). (2001) *Comparative Syntax of the Balkan Languages*. Oxford – New York: Oxford University Press.
- Rosetti, Alexandru (editor/coordinator). (1965) *Istoria limbii române, vol. I*. Bucharest : Editura Academiei Republicii Populare Române.
- Rosetti, Alexandru (editor/coordinator). (1969) *Istoria limbii române, vol. II*. Bucharest : Editura Academiei Republicii Populare Române.
- Rosetti, Alexandru. (1973) *Brève histoire de la langue roumaine des origines à nos jours*. The Hague, Paris: Mouton.
- Rosetti, Alexandru. (1973) *Études Linguistiques*. Paris / The Hague: Mouton.
- Rura, Michael J. (1961) *Reinterpretation of History as a Method of Furthering Communism in Rumania: A Study in Comparative Histiography*. Washington, DC: Georgetown University Press.
- Sandfeld, KR. (1930) *Linguistique balkanique: Problèmes et résultats*. Paris: Librairie Ancienne Honoré Champion.
- Seiver, George O. (1953) *Introduction to Romanian*. New York: Hafner Publishing Company.
- Shafir, Michael. (1985) *Romania: Politics, Economics and Society: Political Stagnation and Simulated Change*. Boulder, CO: Lynne Rienner Publishers, Inc.

- Spolsky, Bernard. (2004) *Language Policy*. Cambridge: Cambridge University Press.
- Ștefănescu-Drăgănești, Virgiliu. (1986) *Romanian Continuity in Roman Dacia: Linguistic Evidence*. Miami Beach: Romanian Historical Studies.
- Swadesh, Morris. (1959) "Linguistics as an Instrument of Prehistory", in: Dell Hymes (editor). *Language in Culture and Society: A Reader in Linguistics and Anthropology*. New York: Harper and Row, pp. 575-584.
- Thomason, Sarah Grey and Terrence Kaufman. (1988) *Language Contact, Creolization, and Genetic Linguistics*. Berkeley / Los Angeles / London: University of California Press.
- Tismancanu, Vladimir. (2003) *Stalinism for All Seasons*. Berkeley: University of California Press.
- Tomić, Olga Mišeska. (2003) "The Balkan Sprachbund Properties", in: Olga Mišeska Tomić (editor). *Topics in Balkan Syntax and Semantics*. Amsterdam: Benjamins, pp. 1-58
- Treptow, Kurt W. and Marcel Popa. (1996) *Historical Dictionary of Romania (European Historical Dictionaries, No. 15)*. Lanham, MD / London: The Scarecrow Press.
- Vincent, Nigel. (1988) "Latin", in: Martin Harris and Nigel Vincent (editors), *The Romance Languages*. New York: Oxford University Press, pp. 26-78.
- Vinogradov, V. V. (editor-in-chief). (1966) *Języki narodów SSSR, tom pierwszy: Indoeuropejskie języki*. Moscow: Izdatel'stvo "Nauka".
- Wikipedia. (2005-2006) "Balkan Linguistic Union," "Romania," "Moldova," "Roman Dacia." Wikimedia Foundation. <<http://www.wikipedia.org/>>

## APPENDIX: ALPHABET AND TRANSLITERATION TABLES

### Romanian / Moldovan Latin Alphabet

<b>Aa</b>	<b>Ăă</b>	<b>Ââ</b>	<b>Bb</b>	<b>Cc</b>	<b>Dd</b>	<b>Ee</b>	<b>Ff</b>	<b>Gg</b>	<b>Hh</b>	<b>Ii</b>	<b>Îî</b>	<b>Jj</b>	<b>Kk</b>
/a/	/ə/	/ɨ/	/b/	/k,č/*	/d/	/e/	/f/	/g,ǵ/*	/h/	/i,y,Ci/**	/ɨ/	/ʒ/	/k/
<b>Ll</b>	<b>Mm</b>	<b>Nn</b>	<b>Oo</b>	<b>Pp</b>	<b>Rr</b>	<b>Ss</b>	<b>Șș</b>	<b>Tt</b>	<b>Țț</b>	<b>Uu</b>	<b>Vv</b>	<b>Xx</b>	<b>Zz</b>
/l/	/m/	/n/	/o/	/p/	/r/	/s/	/š/	/t/	/ts/	/u,w/***	/v/	/ks/	/z/
Diph: <b>ea oa eu</b>													
/eə/ /oa/ /eu/													

\* *C* and *G* are pronounced as /k/ and /g/ before consonants and the vowels *A*, *Ă*, *Â*, *Î*, *O*, and *U*. Before the front vowels *E* and *I*, they are pronounced /č/ and /ǵ/. The sounds /k/ and /g/ do exist in environments preceding front vowels, but they are spelled *CH* and *GH*.

\*\* *I* is typically pronounced as /i/ between consonants. When it precedes a vowel, it is pronounced as the glide /y/. When *I* occurs word finally, it is not pronounced as a vowel, but simply palatalizes the preceding consonant (e.g. *ani* ‘years’ is pronounced as /aɲ/). Word finally, /i/ is spelled *II* (as in *copii* ‘children’). *III* word finally is pronounced /iyi/ (as in *copiii* ‘the children’). The only instances in which a final single *I* is pronounced are when it is stressed as a verb infinitive ending (*a vorbi* ‘to speak’ /a vor`bi/) or when it follows a consonant cluster of the form *Cr* (as in *negri* ‘black pl.’ /`negri/

\*\*\* *U* is pronounced as /w/ before a vowel (as in *ziua* ‘the day’ /ziwa/)

### Moldovan Cyrillic Alphabet

<b>Аа</b>	<b>Бб</b>	<b>Вв</b>	<b>Гг</b>	<b>Дд</b>	<b>Ее*</b>	<b>Жж</b>	<b>Ѕѕ</b>	<b>Зз</b>	<b>Ии</b>	<b>Йй</b>	<b>Кк</b>	<b>Лл</b>	<b>Мм</b>
/a/	/b/	/v/	/g/	/d/	/e,ye/	/ž/	/j/	/z/	/i/	/y/	/k/	/l/	/m/
<b>Нн</b>	<b>Оо</b>	<b>Пп</b>	<b>Рр</b>	<b>Сс</b>	<b>Тт</b>	<b>Уу</b>	<b>Фф</b>	<b>Хх</b>	<b>Цц</b>	<b>Чч</b>	<b>Шш</b>	<b>Ыы</b>	<b>Ьь**</b>
/n/	/o/	/p/	/r/	/s/	/t/	/u/	/f/	/h/	/ts/	/č/	/š/	/ɨ/	/Cj/
<b>Ээ Юю* Яя*</b>													
/ə/ /yu/ /eə,ya/													

\* At the beginning of words and after vowels, the letters *Е*, *Ю*, and *Я* are pronounced as /je/, /ju/, and /ja/. After consonants, *Е* is either pronounced as /e/ or /je/, depending on the word (it corresponds to Latin *E* or *IE*, but does not distinguish between the two orthographically). *Ю* is always pronounced as /ju/. *Я* after consonants is pronounced as /eə/ and sometimes /ja/ (it corresponds to Latin *EA* and *IA*, but does not distinguish between the two orthographically)

\*\* The symbol *Ь* is like the Romanian Latin word-final *I*. It has no sound in and of itself, but palatalizes the preceding consonant (*ani* /aɲ/ from the previous example is spelled *ань* in Cyrillic)

### Russian Cyrillic Alphabet and Transliteration Scheme

<b>Аа</b>	<b>Бб</b>	<b>Вв</b>	<b>Гг</b>	<b>Дд</b>	<b>Ее*</b>	<b>Ёё*</b>	<b>Жж</b>	<b>Зз</b>	<b>Ии</b>	<b>Йй</b>	<b>Кк</b>	<b>Лл</b>	<b>Мм</b>
a	b	v	g	d	e,je	ë,jë	ž	z	i	j	k	l	m
/a/	/b/	/v/	/g/	/d/	/e,je,je/	/o,jo,yo/	/ž/	/z/	/i/	/y/	/k/	/l/	/m/
<b>Нн</b>	<b>Оо</b>	<b>Пп</b>	<b>Рр</b>	<b>Сс</b>	<b>Тт</b>	<b>Уу</b>	<b>Фф</b>	<b>Хх</b>	<b>Цц</b>	<b>Чч</b>	<b>Шш</b>	<b>Щщ</b>	<b>Ъъ**</b>
n	o	p	r	s	t	u	f	h	c	č	š	šč	--
/n/	/o/	/p/	/r/	/s/	/t/	/u/	/f/	/x/	/ts/	/č/	/š/	/šč/	--
<b>Ыы</b>	<b>Ьь***</b>	<b>Ээ</b>	<b>Юю*</b>	<b>Яя*</b>									
y	‘	è	ÿ,ju	ä,ja									
/i/	/ɨ/	/e/	/u,ju,yu/	/ja,ja/									

\* The vowels *Е*, *Ё*, *Ю*, and *Я* are pronounced with the full palatal glide /y/ when they appear word initially or after a vowel. When they occur after a consonant, their usual effect is that they palatalize the preceding consonant. However, certain exceptions do exist: *Е* is pronounced as /e/ when it follows *Ж*, *Ц*, *Ч*, *Ш*, or *Щ*. *Ё* is pronounced as /o/ after *Ж*, *Ч*, *Ш*, or *Щ*, and is never spelled after *Ц* (*О* is used instead). *Ю* is pronounced /u/ when it follows *Ж*, *Ц*, or *Ш* and is never written after *Ч* or *Щ* (*У* is used instead). The transliteration scheme follows mostly a letter-to-letter pattern, except when these letters occur word initially or after vowels, in which case they are transliterated as *JE*, *JË*, *JU*, and *JA*.

\*\* The letter *Ь*, the “hard sign” has only one function: it stops a consonant from being palatalized when it is in an otherwise palatalizing environment, resulting in a consonant cluster *C+/y/*. Its function can be demonstrated by the minimal pair *сесть* ‘to sit’ and *съесть* ‘to eat up (perfect form of *есть* ‘to eat’), which are pronounced /sʲestʲ/ and /sjestʲ/, respectively. The transliteration ignores this letter, transliterating the words as *sest*’ and *sjest*’.

\*\*\* The letter *Ь*, the “soft sign” palatalizes the preceding consonant (as in *семь* ‘seven’, pronounced /sʲemʲ/. This symbol has no effect on the letters *Ж*, *Ч*, *Ш*, or *Щ*, though it may often be spelled after them for historical and grammatical reasons. It is transliterated as an apostrophe.



## **VITA**

Chase Faucheux, originally of Arabi, Louisiana, now calls Folsom, Louisiana his home. After graduating from St. Paul's School in Covington, Louisiana in 2000, he attended Tulane University in New Orleans. He graduated from Tulane with a Bachelor of Arts degree in the spring of 2004, majoring in Linguistics and minoring in Spanish. It was here that his interest in linguistics was formed, specifically in the areas of historical linguistics and language and politics. At Tulane, he studied Spanish, Russian, and Arabic in addition to his linguistics curriculum. Mr. Faucheux entered the Louisiana State University Interdepartmental Graduate Program in Linguistics in the fall of 2004, and will graduate in the summer of 2006. At Louisiana State University, he worked as an English as a Second Language instructor for the English Language Orientation Program, where he prepared international students for further study in American universities. Mr. Faucheux will move to Baltimore, Maryland in the fall of 2006, where he will work as a linguist.